ENSEMBLE STATISTICS AND ERROR COVARIANCE IN A RAPIDLY INTENSIFYING HURRICANE

A Thesis

by

MATTHEW CHARLES RIGNEY

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

May 2009

Major Subject: Atmospheric Science

ENSEMBLE STATISTICS AND ERROR COVARIANCE IN A RAPIDLY INTENSIFYING HURRICANE

A Thesis

by

MATTHEW CHARLES RIGNEY

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Approved by:

Co-Chairs of Committee,	Fuqing Zhang
	Kenneth Bowman
Committee Members,	Mikyoung Jun
	R. Saravanan
Head of Department,	Kenneth Bowman

May 2009

Major Subject: Atmospheric Science

ABSTRACT

Ensemble Statistics and Error Covariance in a Rapidly Intensifying Hurricane. (May 2009)

Matthew Charles Rigney, B.S., Rice University

Co-Chairs of Advisory Committee: Dr. Fuqing Zhang
Dr. Kenneth Bowman

This thesis presents an investigation of ensemble Gaussianity, the effect of non-Gaussianity on covariance structures, storm-centered data assimilation techniques, and the relationship between commonly used data assimilation variables and the underlying dynamics for the case of Hurricane Humberto. Using an Ensemble Kalman Filter (EnKF), a comparison of data assimilation results in Storm-centered and Eulerian coordinate systems is made. In addition, the extent of the non-Gaussianity of the model ensemble is investigated and quantified. The effect of this non-Gaussianity on covariance structures, which play an integral role in the EnKF data assimilation scheme, is then explored. Finally, the correlation structures calculated from a Weather Research Forecast (WRF) ensemble forecast of several state variables are investigated in order to better understand the dynamics of this rapidly intensifying cyclone.

Hurricane Humberto rapidly intensified in the northwestern Gulf of Mexico from a tropical disturbance to a strong category one hurricane with 90 mph winds in 24 hours. Numerical models did not capture the intensification of Humberto well. This could be due in large part to initial condition error, which can be addressed by data assimilation

schemes. Because the EnKF scheme is a linear theory developed on the assumption of the normality of the ensemble distribution, non-Gaussianity in the ensemble distribution used could affect the EnKF update. It is shown that multiple state variables do indeed show significant non-Gaussianity through an inspection of statistical moments.

In addition, storm-centered data assimilation schemes present an alternative to traditional Eulerian schemes by emphasizing the centrality of the cyclone to the assimilation window. This allows for an update that is most effective in the vicinity of the storm center, which is of most concern in mesoscale events such as Humberto.

Finally, the effect of non-Gaussian distributions on covariance structures is examined through data transformations of normal distributions. Various standard transformations of two Gaussian distributions are made. Skewness, kurtosis, and correlation between the two distributions are taken before and after the transformations. It can be seen that there is a relationship between a change in skewness and kurtosis and the correlation between the distributions. These effects are then taken into consideration as the dynamics contributing to the rapid intensification of Humberto are explored through correlation structures.

DEDICATION

I would like to dedicate this manuscript to my dad, who introduced me to and cultured my love of weather, my mom who spent years teaching me how to think, and to my wife, a source of constant strength and encouragement.

ACKNOWLEDGEMENTS

I would first and foremost like to thank my advisors, Dr. Fuqing Zhang and Dr. Ken Bowman for the support that they have provided me in my time at Texas A&M University. Over the last two and a half years, I have been pushed and stretched to new limits. Although at times I resented the demands placed upon me, because of them, I am a better thinker and a better scientist. In addition, I would like to thank my other committee members for their help in completing and editing this manuscript. My family and friends have also played an integral role in the completion of my thesis. Without their love and support I would not be where I am today. Even though they have always stood beside me, I would like to thank them even more for pushing me and encouraging me. To my mom and dad, I would like to say "thank you" for the years of love showered upon me and the investment you made in my education. Even when I moved away, you reminded me to live by the same principles that you worked many years to instill in my character.

TABLE OF CONTENTS

Page

ABSTRACT iii

DEDICATION v

ACKNOWLEDGEMENTS vi

TABLE OF CONTENTS vii

LIST OF FIGURES ix

LIST OF TABLES x

1. INTRODUCTION. 1

2. ENSEMBLE STATISTICS OF A LORENZ SYSTEM 8

3. ENSEMBLE STATISTICS OF A TROPICAL CYCLONE 14

    3.1 Synoptic Setup 14
    3.2 The Forecast Model 15
    3.3 Ensemble Statistics 16
    3.4 Area Averaging ("Bundling") 22
    3.5 Non-Gaussianity's Effect on Correlation Structures 23

4. ERROR COVARIANCE STRUCTURES 25

5. IMPACT OF STORM-CENTERED ASSIMILATION SCHEME 29

6. SUMMARY AND CONCLUSIONS 34

REFERENCES 59

APPENDIX A 39

VITA 59

LIST OF FIGURES

FIGURE		Page
2.1	Sample ensemble distribution of L95 Lorenz model	39
2.2	Sample distributions of unimodal distributions	40
2.3	Ensemble mean and standard deviation of L95 Lorenz model	41
2.4	Ensemble skewness and kurtosis of L95 Lorenz model	42
3.1	Configuration of the WRF model domains 1, 2, and 3	44
3.2	Ensemble mean mixing ratio, skewness, and kurtosis	45
3.3	Ensemble mean zonal winds, skewness, and kurtosis	46
3.4	Ensemble mean meridional winds, skewness, and kurtosis	47
3.5	Ensemble mean pressure, skewness, and kurtosis	48
3.6	Effect of area averaging on Gaussianity of pressure	49
3.7	Effect of area averaging on Gaussianity of mixing ratio	50
3.8	Effect of area averaging on Gaussianity of vertical velocity	51
4.1	Correlation of 850 hPa potential vorticity with surface mixing ratio and surface temperature	53
4.2	Correlation of 850 hPa potential vorticity with 200 hPa zonal wind and meridional wind	54
5.1	Pressure field of ensemble member 11	55
5.2	Ensemble mean pressure in storm-centered and Eulerian frames	56
5.3	Storm-centered coordinate system and Eulerian coordinate system analyses	57

FIGURE Page

5.4 Normalized measure of error reduction..... 58

LIST OF TABLES

TABLE Page

2.1 Effect of ensemble size on Gaussianity..... 43

3.1 Effect of non-Gaussian transformations on correlation structures 52

1. INTRODUCTION

Although there have been improvements in forecasting the intensity and genesis of mesoscale events (Hawblitzel 2007), accurately forecasting the genesis and evolution of tropical cyclones remains challenging. Data assimilation, one of the most important sub-fields in numerical weather prediction, holds one key in addressing this predictability problem. This sub-field of numerical weather prediction encompasses efforts to combine all available data to optimally estimate the state of the atmosphere. Because the evolution of weather systems is highly sensitive to the accuracy of the initial conditions of the system, improving the estimation of the initial atmospheric state is vital. This sensitivity is especially present in tropical systems. Although track-forecasting skill has drastically increased in recent years (Elsberry 2007), and 48-hour track forecasts are now as accurate as 24-hour forecasts from 10 years ago (Franklin 2005), Hurricane Humberto (2007), along with many other tropical cyclones over the past decade, show that intensity forecasts have not made the same progress (Franklin 2005).

None of the initial models used to forecast Humberto (2007) predicted the cyclone to become a hurricane, and in fact, most kept the system at minimal tropical cyclone strength. Evaluating data assimilation schemes provide one way of continuing to improve our ability to forecast these complex systems. Previous studies of Hurricane Humberto have shown that the use of an Ensemble Kalman filter data assimilation scheme assimilating Doppler radar observations produces forecasts superior to those

---

This thesis follows the style of *Monthly Weather Review*.

used operationally (Zhang, et al. 2009). In this study, new applications of the Ensemble Kalman filter will be explored and ensemble statistics used in the Ensemble Kalman filter scheme will be examined to determine the suitability of the scheme in areas with nonlinear dynamics.

Several data assimilation schemes have been used operationally, but this manuscript will focus on the promising EnKF scheme. The 3D-variational method (3DVAR), the most prominently used operational method, uses a cost function to optimally minimize error. It also requires an adjoint model to operate. More recently, the 4D-variational method (4DVAR) has been developed as an extension of the 3DVAR. In the 4DVAR, the cost function is also minimized with respect to time. In contrast, the EnKF is a sequential data assimilation scheme. As observations are made, they are sequentially fed into the assimilation scheme and assimilated according to the equations developed to maximally reduce error. Sequential data assimilation has shown promise in multiple situations from synoptic scale to mesoscale (Zhang 2005, Hawblitzel et al., 2007).

Most operational data assimilation schemes such as the 3DVAR use isotropic, stationary background error covariance structures. The EnKF and ensemble forecasting provide a means to garner essential information in estimating flow-dependent background error covariance and applying that information to improve data assimilation schemes. Numerous studies, from idealized examples to full-scale assimilation of real observations have shown promising results for ensemble based assimilation schemes. Until recently, however, the evolution of these flow-dependent background error

covariance structures had not been closely studied. Zhang (2005) used an ensemble to demonstrate the anisotropic growth of background error covariance structures in the case of a rapidly intensifying winter storm over the mid-Atlantic region. In simulations of the event, initial, random perturbations quickly evolved into coherent structures with spatial correlations between the same and different forecast variables. Further investigation showed the strongest coherence and highest correlations of this cross covariance (covariance/correlation between different state variables) in areas with high potential vorticity gradients, active moist dynamics, or over regions of strong cyclogenesis. Ultimately, these correlation structures were shown to be anisotropic and their evolution governed by the underlying dynamics of the system. Thus, EnKF schemes, governed by dynamic correlation structures allow data assimilation schemes to evolve with the system.

One other field of recent interest involves comparing the assimilation of observations in a Storm-centered versus Eulerian assimilation window. Assimilation done in a Storm-centered scheme is object oriented (Lawson and Hansen, 2005), whereas the Eulerian assimilation window is geographically oriented. A Storm-centered assimilation system could prove especially useful for smaller mesoscale systems. In the case of a hurricane, assimilating in an object-oriented frame could improve calculations of ensemble based covariance structures since all storms and relevant features of storms are centered at the same place.

Tempering enthusiasm about the EnKF, ensemble non-Gaussianity could

potentially lead to a degradation of the EnKF scheme's results. Two major assumptions are made when developing and using the EnKF. First, it is assumed that all errors are Gaussian and secondly, that all error structures evolve linearly. Most atmospheric systems include local areas of nonlinear evolution, and with unlimited computing power, the extended Kalman filter could be implemented to perfectly solve for nonlinear systems (Jazwinski, 1970). However, today's resources are limited, and instead, adaptations must be made to the EnKF. Lawson and Hanson (2005) found that the EnKF could be adapted to find a minimum variance solution of the EnKF in nonlinear cases. This is not nearly as useful of a solution as the most probable solution. Although adaptations have been attempted, they have generally been on simple, lower-dimensional systems. The severity and extent of non-Gaussianity still must be explored for real atmospheric systems.

The sequential EnKF scheme was first proposed by Evensen (1994) as an approximation of the computationally demanding extended Kalman filter, first proposed by Kalman (1960). The EnKF estimates the state of the atmosphere by combining a prior or short term forecast with a set of observations. With each new set of observations that is fed into the assimilation scheme, the state of the atmosphere is updated by an amount proportional to the difference between the forecast at the location of the observation and the observation. This difference is known as the innovation. The relationship between the true state of the atmosphere, the forecast (or background), and the innovation can be written as:

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}[\mathbf{y}_o - \mathbf{H}(\mathbf{x}_b)] \quad (1.1).$$

In (1.1), $\mathbf{x}_a$ is the analysis state vector (i.e. the best estimate of the true state of the atmosphere), $\mathbf{x}_b$ is the "first guess" or prior, $\mathbf{y}_o$ is the observation vector, and $\mathbf{H}$ is a transformation matrix that maps the prior forecast or "first guess" to observation space through a chosen interpolation scheme.

$\mathbf{K}$ is the Kalman gain matrix. This gain matrix controls the amount by which the background forecast is changed by each set of observations in order to minimize the analysis error. In the EnKF, this matrix is given by

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1} \quad (1.2).$$

and the analysis error is given by

$$\mathbf{P}_a^{-1} = \mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} \quad (1.3)$$

$\mathbf{B}$ is the background covariance error matrix and $\mathbf{R}$ is the observation error covariance. Thus, the value of the Kalman gain matrix is controlled completely by the ratio of the forecast uncertainty to the total uncertainty (forecast uncertainty plus observational uncertainty). Because the gain matrix is intimately tied to the background covariance error, and background covariance error has been shown to be flow-dependent and dependent on the dynamics of a system (Zhang 2005), a data assimilation scheme that does not have a time-varying background error covariance will be inferior in calculating the analysis of the atmospheric state. This gain matrix will fail to capture errors brought about by the dynamical evolution of the system.

Unlike other data assimilation schemes such as the 3-D variational method, which has an isotropic, and essentially time-invariant background error covariance (Houtekamer and Mitchell, 1998), the EnKF allows for a flow-dependent background covariance through its ensemble nature. In an EnKF scheme, ensemble spread is used to estimate the background covariance structures by

$$\frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})(y_i-\bar{y}) \qquad (1.4).$$

In (1.4), x and y are state variables, $\bar{x}$ and $\bar{y}$ are the ensemble means of the state variables, and N is the number of ensemble members. Since the background covariance error is based upon the ensemble, it is then also based solely upon the dynamics of the ensemble and is completely flow dependent and non-isotropic. This allows, in theory, for better estimation of the Kalman gain matrix, and in turn a better representation of the true state of the atmosphere.

In previous studies, the EnKF has performed well. Zhang et al. (2005) used an EnKF scheme in a study of the "surprise" snowstorm of 2000. Although there were limitations in the study, such as a perfect model assumption, the EnKF scheme significantly reduced (60-80%) the large-scale error when compared to an ensemble with no data assimilation scheme. In addition, the EnKF proved to remain effective under adverse circumstances, such as large observational errors or complete loss of sounding data. Cases with "imperfect models" have also shown promising results. Investigating the same "surprise" snowstorm, Meng and Zhang (2007) showed a 36-67% reduction in

error. Other studies at larger, synoptic scales have also shown promising performance for the EnKF (Houtekamer et al. 2005).

Hurricane Humberto will be used as the test bed for this study. As stated previously, data assimilation is of key importance for predicting the evolution of systems in which there is potential for rapid intensification or intensification is predicated on warm, moist dynamics. Humberto, one of the most rapidly intensifying tropical cyclones ever recorded (Blake 2007) fits into both of these categories, and thus is an ideal system to study new data assimilation methods and ensemble characteristics.

2. ENSEMBLE STATISTICS OF A LORENZ SYSTEM

Evensen (1994) developed the EnKF system based upon assumptions of ensemble linearity and normality, following the rules of Gaussian statistics. Currently, statistical characteristics of ensemble distributions used in EnKF schemes have not been thoroughly investigated. Before examining a real world numerical weather prediction simulation, however, it will be beneficial to inspect a simpler case to lay the expectations and framework for a more in depth study. By evaluating the statistical characteristics of an 80-variable, 10,000-member ensemble, the groundwork for a larger study using a operational Weather Research and Forecasting (WRF) model can be laid.

The L95 Lorenz model is an 80 variable Lorenz model. At time t_0 in this study, the model's basic state value is between four and five. Small, random perturbations are added to the initial state of all 80 variables to initialize the ensemble. The model is then integrated forward in time for 14,400 time steps or approximately 10 years with no data assimilation. After integration, the model output can be systematically examined to determine the time evolution characteristics of the L95 Lorenz model.

First, the distribution of several variables is visually inspected. The ensemble distributions of two particular variables are plotted as a histogram with a Gaussian overlay (Fig 2.1). Although the distributions appears mostly Gaussian, it should be noted that it is actually slightly flatter at the top of the distribution and has more restricted tails than a Gaussian distribution. In addition, it is also slightly skewed. This histogram can be referenced to sample distributions (Fig 2.2), which are simple

distributions plotted for various values of kurtosis. This distribution is representative of many points in the Lorenz model, and it is for distributions such as these that the effect of the non-Gaussianity must be determined.

Next, the ensemble mean of the model (Fig. 2.3) is examined. The ensemble mean is calculated by averaging the value of all ensemble members for each of the 80 variables in the model. It is noted that the values for the 80 variables range from a minimum of approximately 2.25 to a maximum of approximately 2.43. This range of values is below the basic state of four to five, indicating that, the Lorenz model has more strongly diverged in a negative direction in the absence of data assimilation.

To determine the amount of spread among ensemble members, the standard deviation of each variable is studied (Fig. 2.3). Standard deviation indicates the spread of a variable about the mean. Larger standard deviations indicate a distribution higher dispersion around the mean value. The range of standard deviation is approximately 3.59 to 3.69. Given the ensemble mean ranged from 2.25 to 2.43, and the original basic state value of the model was four meaning that the standard deviation is almost as great as the basic state value of the variable and larger than the ensemble mean after integration forward in time, indicating high ensemble spread. This high spread is due to two factors. First, no data assimilation was performed throughout the integration, and thus each variable was allowed to freely diverge from its initial state. Secondly, a Lorenz model has multiple attractor basins. As each variable with a slightly different initial condition wanders towards distinct attractors, the spread between variables increases.

Skewness (Fig. 2.4) is the third statistical moment and indicates whether a distribution is "leaning" to the right or left in comparison to a Gaussian distribution. It is the first statistical moment that gives information regarding the Gaussianity of a distribution. For the L95 model in this study, the skewness ranges from 0.0497 to 0.1341. Positive values of skewness point to a right skewed distribution. Being right skewed means that the distribution, when compared to a Gaussian distribution, has a heavier tail to the right and a shorter, less significant tail on the left side of the distribution. This is the first sign that the distribution in the L95 Lorenz model could be non-Gaussian in nature.

Finally, the fourth statistical moment, kurtosis (Fig 2.4), indicates the amount of peakedness in a distribution. A value of zero indicates a Gaussian function. For this L95 Lorenz model, the values of kurtosis range from -0.5716 to -0.4640. These values show a platykurtic distribution, meaning that there is less peakedness than in a Gaussian distribution and the distribution has a peak more like a semi-circle compared to a Gaussian distribution.

Synthesizing the data on the 10,000-member L95 Lorenz model shows that there are elements of non-Gaussianity in the ensemble distribution. Values of kurtosis with magnitude as high as 0.5716 indicate levels of non-Gaussianity diverging significantly from the normal. Although a 10,000-member, L95 Lorenz model does not necessarily behave like an operational primitive equation model, these initial results motivate the continuation of this study.

Determination of ensemble size has been largely empirical and dictated by the need for an ensemble with adequate spread and an ensemble that is computationally efficient and feasible. Since operational ensembles cannot have 10,000 members due to computational costs, it will be beneficial to explore the extent that ensemble size affects the ensemble distribution statistics. If decreased ensemble size increases non-Gaussianity, problems with an EnKF scheme could occur. In most of today's ensemble based forecasting, an ensemble size of between 20 and 50 members is used. Because the L95 Lorenz model is efficient and computationally quick, it can be used to indicate potential behavior of primitive equation numerical models' ensemble distributions for ensemble sizes that are small when compared to many ensembles used in the statistics field. Determining the optimal size of an ensemble for computational cost while maintaining key statistical features is integral if ensemble forecasting is to continue to evolve and improve.

To examine the effect of size on ensemble statistics, 30 ensemble members are randomly selected and ensemble statistics calculated. These 30 members were selected using MATLAB's randn() function to generate 30 random numbers between 1 and 10,000. The numbers generated were the ensemble members selected for the 30-member ensemble. Table 2.1 shows values of the first through fourth statistical moments for both ensembles. For each moment, the maximum, minimum, mean, and standard deviation is noted. This gives a complete description of the data for both ensembles and each moment.

Comparing the ensemble mean shows that the mean of the ensemble mean is almost identical in the two ensembles. In fact, the mean of the ensemble mean of the 30-member ensemble is only 1.7% higher than that of the 10,000-member ensemble. Although the mean is approximately the same for both ensembles, ensemble mean is not overly important to the Gaussianity of the EnKF. It does, however, show the distributions will look similar in their prediction of events, even if the underlying distributions turn out to differ.

Comparing skewness and kurtosis of the two ensembles more directly addresses the amount of non-Gaussianity in the distribution and the potential impact of ensemble size on the EnKF scheme. Skewness, the third statistical moment, of the 30-member ensemble has a mean skewness value of 0.1357. It was 0.0940 in the 10,000-member ensemble, a 44.3% increase. In addition, the range of values for the 30-member ensemble is much higher than the 10,000-member ensemble. It is hard to make definitive conclusions about the skewness based on this data because the standard deviation of the 30-member ensemble is high enough to encompass the mean of the 10,000-member ensemble, but the smaller ensemble size shows marked increase in skewness. Average kurtosis increases to 3.4395, a 655% increase in magnitude and a change in sign. This indicates that reducing ensemble size not only can increase the kurtosis of a distribution, but it can also completely change the type of non-Gaussianity present in a distribution. This portion of the study gives evidence that non-Gaussianity is introduced into an ensemble through both the dynamical evolution of the system and the inherent inability of a small ensemble to capture all possible evolutions of a system.

Studying the 10,000-member L95 Lorenz model has yielded preliminary information on ensemble distributions in atmospheric science models. It is determined that the L95 Lorenz model exhibits non-Gaussian characteristics in both a 10,000-member ensemble and a 30-member ensemble. The non-Gaussianity noted is greater in 30-member ensemble. This is especially true of the skewness and kurtosis values, which increase significantly in the smaller ensemble. Because the simple L95 Lorenz model shows evidence of non-Gaussianity, motivation exists to characterize the non-Gaussianity present in a numerical weather simulation of a system involving intense convection and non-linear processes.

3. ENSEMBLE STATISTICS OF A TROPICAL CYCLONE

3.1 Synoptic Setup

On September 5, a frontal boundary sagged south off the coast of Florida into the southeast Gulf of Mexico. Over the next several days, this trough moved to the west-northwest in response to ridging over the southeastern United States. By September 11, convection began to fire southeast of Galveston and that night a surface low began to form. At 0900 UTC on September 12, 2007, the National Hurricane Center in Miami, Florida classified the system as a tropical depression (Blake 2007).

Within three hours of the initial advisory, the National Hurricane Center upgraded the tropical depression to tropical storm Humberto. As the system drifted slowly northward towards the Texas coast, well-defined banding features became more prominent on radar from Houston and the rapid intensification of Humberto continued throughout the day. As evening approached, the storm turned to the north-northeast around the western periphery of the ridge and took aim at the Texas/Louisiana state line.

By the early morning hours of September 13, 2007, radar and reconnaissance indicated that the storm had reached hurricane intensity. At 0400 UTC 13 September, Humberto reached its peak intensity of 80 kts about 20 miles south of High Island, Texas (Blake 2007). About three hours later, Humberto made landfall just east of High Island, Texas, with a minimum pressure of 985 millibars. It then quickly weakened over

the next 12-24 hours before finally dissipating as it moved northeastward over Louisiana and Mississippi.

None of the major operational models (GFDL, UKMET, ECMWF, and NOGAPS) forecasted the rapid intensification of Humberto, and this was reflected in the official forecast from the National Hurricane Center. In one of the first advisories issued on Humberto at 1500 UTC on September 12, less than 24 hours before peak intensity, the National Hurricane Center predicted Humberto to reach a peak intensity of 40 kt. Twelve hours later, when Humberto had reached an intensity of 55 kt and was only 4 hours from landfall, the National Hurricane Center predicted that "some additional strengthening could occur before the center makes landfall...and winds could be approaching hurricane force" (NHC Forecast Discussion, 2007), however, it was not explicitly forecast to become a hurricane and no hurricane warnings or watches were issued at the time. Ensemble forecasting is intended to capture some of the variability inherent in forecasting intense weather systems and because of the poor performance of deterministic models, Humberto makes an attractive case study on the use of ensemble forecasting and the effect of data assimilation in tropical cyclone forecasting.

3.2 The Forecast Model

The advanced WRF (ARW) is used for this study. WRF is a fully compressible, nonhydrostatic mesoscale model (Skamarock et al. 2005). The vertical coordinate follows the terrain using hydrostatic pressure, and the model uses an Arakawa-C grid.

Prognostic variables are column mass of dry air, velocities (u, v, w), potential temperature, geopotential, and mixing ratio for water vapor, cloud, rain, ice, snow, and graupel.

In the control experiments, three model domains with two-way nesting are used. The coarse domain covers the contiguous United States with 160x121 grid points and a grid spacing of 40.5 km, and the inner domain D2 (D3) cover the central United States with 160x121 (253x253) grid points and a grid spacing of 13.5 (4.5) km (Fig. 3.1). All model domains have 35 vertical layers, and the model top is set at 10 hPa. The physical parameterization schemes include the Grell-Deveni cumulus scheme (Grell and Devenyi 2002), WRF Single Moment 6-class microphysics with graupel (Hong et al. 2004) and the Yonsei University (YSU) scheme (Noh et al. 2003) for planetary boundary layer processes. The NCEP GFS operational analysis at 00Z 12 September and its forecast are used to create initial and boundary conditions. Data assimilation is performed for all domains, but all verification is performed for D3. Ensemble size was set to 30 members in this study. In previous studies (e.g., Houtekamer and Mitchell 2001, Anderson 2001; Snyder and Zhang 2003; Zhang 2005; Zhang et al. 2006; Meng and Zhang 2007; Meng and Zhang 2008a,b), an ensemble size of 20-50 was found to be both reasonable and affordable.

3.3 Ensemble Statistics

WRF ensemble statistics will be explored using the same methodology as with the L95 model in the previous section. These calculations are important to determine if

the linear assumptions upon which the EnKF is based are valid. By calculating the first through fourth statistical moments of the ensemble, the Gaussianity of the ensemble can be analyzed. Because the EnKF scheme is based upon Gaussian statistics, any significant non-Gaussianity of the ensemble distribution will produce less than optimal results. Quantifying the effects of non-Gaussianity will also be a major component of this section.

To analyze the ensemble characteristics of the WRF model run, the first through fourth statistical moments across the entire model domain will be calculated at each grid point at 0900 UTC on 13 September when Humberto was at peak intensity near landfall. In this portion of the study, statistical moments will be averaged spatially and at each vertical level for several state variables important to EnKF data assimilation to determine the Gaussianity across the entire domain. Variables that show evidence of high non-Gaussianity will be looked at more closely by re-calculating statistics over a smaller domain size centered on the storm in order to investigate the role that convection plays in producing non-Gaussianity in the WRF ensemble.

One of the primary variables assimilated in many data assimilation schemes is mixing ratio (Fig. 3.2) since one of the primary areas of error growth is in locations with moist dynamics (Zhang, 2005). First it should be noted that the average mixing ratio decreases rapidly with height, becoming negligible around 12 km. In addition, as domain size becomes more tightly focused on the storm, mixing ratio increases at the surface from 14 g/kg when averaged over the entire domain to just over 19 g/kg in a 20x20 gridpoint sub-domain centered over Humberto. Skewness reaches a maximum in

the large domain in the mid-levels, where values reach almost 1.1. As the domain size decreases, meaning that a larger percentage of the domain is convective in nature, the skewness decreases across most levels. In fact, for the smallest domain (21x21 gridpoints), skewness reaches a maximum value of 0.7 at 4 km and then increases rapidly above 8 km as the amount of moisture in the atmosphere quickly decreases. Finally, kurtosis is found to greatest when averaged over the entire domain. In the case of averaging over the entire 252x252 gridpoint domain, the kurtosis reaches a maximum value of 2.5 between 2 km and 4 km in height. When averaged over smaller domain sizes, the kurtosis decreases to values of approximately 1.0 at a height of 5 km. This corresponds well to the high values of skewness at approximately the same height. In conjunction with each other, it can be concluded that the ensemble distribution of mixing ratio exhibits significant non-Gaussianity in the mid-levels of the atmosphere. Interestingly, values of skewness and kurtosis are actually larger over larger domains. This is somewhat counter-intuitive since we expect non-Gaussianity to increase with increasing presence of non-linear dynamics. However, since mixing ratio is extremely low at many points in the larger domain, we could expect areas with large amounts of non-Gaussianiy in the larger domains.

Zonal (Fig. 3.3) and meridional (Fig. 3.4) winds also play an integral role in data assimilation schemes. Because horizontal winds are one of the easiest of all state variables to observe and are explicitly calculated within the model, they are one of the most data rich and important variables to assimilate. In the data from the WRF model for Hurricane Humberto, both meridional and zonal averaged winds increase in

magnitude as the domain size decreases. Winds also decrease to near zero in the upper troposphere in the smaller domain size, indicating high pressure aloft. This setup is conducive for strengthening. Skewness values remain relatively constant and have similar values for both meridional and zonal winds. Skewness ranges from about 0.5 to 0.9 near the surface and then remains fairly constant with height, decreasing slightly to approximately 0.5. At the surface, smaller domains have slightly larger skewness values.

Finally, kurtosis values range from 0.8 (zonal winds) to 1.8 (meridional winds) and remain constant with height. All three domain sizes exhibit comparable amounts of kurtosis at all vertical levels. Compared to mixing ratio, zonal and meridional winds have moderate degrees of non-Gaussianity. However, non-Gaussianity is still present and possibly significant to the execution of an EnKF scheme.

Pressure is another measure of tropical cyclone strength. Since it is observed frequently, it is a useful variable in assimilation schemes. In this study, skewness and kurtosis of pressure was found to have significant differences between larger domains and smaller domains (Fig. 3.5). As expected, pressure is slightly lower when averaged over successively smaller domains. In addition, as domain size is decreased, standard deviation increases, especially near the surface. This is expected since the pressure gradient is strong near the center of the storm. Both skewness and kurtosis of pressure paint interesting pictures of convection's role in the Gaussianity of the ensemble distribution. Looking first at skewness, the values averaged over the entire 252x252 gridpoint domain range from 0.4 at the surface, increases in the mid-levels, and then

decreases again in the upper levels. Although these values are moderate skewness values, they pale when compared to the values when averaged over smaller, storm-centered domains. In fact, when averaged over the 50x50 gridpoint domain, the skewness at the surface increase over 100% in magnitude to 0.85. The magnitude again increases to 1.15 when averaged over the 20x20 domain. This indicates significant skewness in the distribution of pressure when measured around the ensemble mean center of the storm and is the first strong evidence that convection plays a key role in significant non-Gaussianity.

This trend continues in the kurtosis of pressure. At the surface, the values of kurtosis in the large domain are around 0.9 in the lowest kilometer of the atmosphere. Like skewness, however, this value increases to 1.5 when averaged over the 50x50 gridpoint domain and to 2.2 when averaged over the 20x20 gridpoint domain. Although the difference between the kurtosis values at different domain sizes decreases with height, pressure at the surface is the most relevant measurement to storm strength. Thus, the distribution of pressure is non-Gaussian near the surface, and highly non-Gaussian in areas of intense convection. This could be due in part to the non-linear nature of convective dynamics. Even more importantly, however, it raises questions about the use of a data assimilation scheme based on linear statistics.

The assumption of Gaussian ensemble distributions has been shown to be a good approximation in many cases at many vertical levels, such as pressure in the mid and upper levels and meridional and zonal winds. This assumption has been shown to be even more accurate when the ensemble domain is largely non-convective. However,

some cases, highlighted most strongly by surface pressure, show strong non-Gaussian behavior in areas of intense convection. What affect this non-Gaussian distribution has on correlation structures and in turn the EnKF scheme then becomes a key question.

3.4 Area Averaging ("Bundling")

Area averaging (bundling) was performed on several state variables to observe its affect on the non-Gaussianity of the distributions. An observed decrease in skewness and kurtosis would advocate for the usefulness of "super-obbing". Super-obbing is the practice of taking an average, or subset, of observations (such as radial velocities from a radar) and incorporating them into one observation used as representative of the entire area. In this investigation, averages are taken over a 2x2 grid point area, 4x4 grid point area, and an 8x8 grid point area. Examining the skewness and kurtosis of pressure (Fig. 3.6), there are only negligible changes in skewness and kurtosis. However, when the domain size is reduced to a 51x51 grid point storm-centered area, the 4x4 area averages begin to show reduced skewness and kurtosis. When the domain size is reduced even further to a 21x21 grid point, storm-centered box, skewness and kurtosis is even more noticeably reduced. In fact, at low to mid-levels, skewness is reduced by over 10% and kurtosis by almost 15%.

Similar findings, although not as uniform, can be found when inspecting the results for mixing ratio (Fig. 3.7). In this particular case, however, the skewness and kurtosis are generally reduced at all levels in the 252x252 domain, while at the smaller domain sizes, the improvement is not evident at all vertical levels. Although these results are not as persuasive as the findings for pressure, there is still evidence present that area-

averaging can improve Gaussianity.

Finally, vertical velocity shows the most potential improvement by bundling. Figure 3.8 shows clear decreases in kurtosis at all domain sizes. Reductions at all domain sizes and most vertical levels are greater than 50% for the 4x4 grid point averages. Skewness also shows a decrease at almost every vertical level and every domain size. The evidence here is most conclusive out of all of the variables examined.

It has been shown that across multiple variables, area-averaging increases Gaussianity. In the most significant cases, kurtosis is reduced by more than half. This lends support to the practice of super-obbing in order to create more normal ensemble distributions. By creating and using super-obs in a EnKF scheme, the potentially more normal distribution has the ability to perform better under the linear assumptions imposed by the EnKF.

3.5 Non-Gaussianity's Effect on Correlation Structures

The effect of non-Gaussianity on correlation structures will be explored through an empirical investigation of correlation structures from Gaussian distributions put through data transformations. Ideally, each non-Gaussian distribution from the WRF Humberto run would be analyzed to investigate the effect on that particular correlation structure. However, the statistics involved would be highly complex and beyond the scope of this study. Instead, known Gaussian distributions will be subjected to data transformations. Correlations will be calculated between two Gaussian distributions and two transformed distributions. In addition, the correlation between a Gaussian distribution, "A," and a post-transformation distribution, "B," will be calculated. By examining the effect on

correlation that these data transformations have, a reasonable inference can be made on whether non-Gaussianity affects the results of the EnKF data assimilation scheme.

Table 3.1 details the results of this study. In addition to correlation values, skewness and kurtosis values of each distribution are also recorded. The original Gaussian distributions are transformed with x^2, x^3, and $\frac{1}{x}$ transformations. For transformed data in Table 3.1, "Correlation 1" refers to the correlation between one Gaussian distribution and one transformed distribution and "Correlation 2" refers to the correlation between two transformed distributions. One clear result of this limited study is that correlation structures were changed significantly only when the mean of the distribution was near zero. In the x^2 and x^3 case, however, the lower correlation was only noted when one transformed Gaussian distribution was correlated with a Gaussian distribution. In the $\frac{1}{x}$ transformation, two transformed, non-Gaussian distributions and one transformed distribution paired with a Gaussian distribution both showed almost no correlation.

Interestingly, the other two cases (mean = 270 and mean = -15) show little decrease in correlation values after the transformation. However, when comparing the values of skewness and kurtosis for these transformations, it is noticed that they are much lower than in the first case with mean zero. In fact, there is no appreciable increase in the non-Gaussianity of the post-transformation distributions for the case with mean set to 270. In addition, for the case where mean is set to -15, skewness and kurtosis increase after the data transformation, but levels are still generally below one

which may not be significant enough to impact correlation structures in this particular case.

This simple study shows that for small increases in non-Gaussianity, correlation is not greatly affected (see cases where mean = 270 and mean = -15). However, as non-Gaussianity increases greatly, correlation structures change greatly, as in the case where the mean is zero. While this study is preliminary, it shows that excessive non-Gaussianity in ensemble distributions can potentially affect correlation structures of those ensembles. These results should motivate further study in this area to detail the specific effect that ensemble non-Gaussianity has on correlation structures used in data assimilation schemes.

4. ERROR COVARIANCE STRUCTURES

Covariance structures of state variables can provide key information in understanding the dynamical setup of Humberto. Why Hurricane Humberto strengthened as rapidly as it did as it approached the upper Texas coast remains a vexing question, especially in the face of a model consensus unanimously opposed to the idea. Ultimately, knowledge of the dynamical setup surrounding the storm will lead to a better understanding of what environmental conditions contribute to intensification. In addition, if variables important to data assimilation also play a key role in the dynamics and intensification of Humberto, further weight can be given to the value of an EnKF scheme. In this section, correlation structures will be used to analyze the major factors contributing to the rapid strengthening of Humberto. Vorticity at 850-hPa is used as a proxy for storm strength. Correlations of 850-hPa potential vorticity averaged over a 10 gridpoint square surrounding the storm center with several state variables are looked at to determine what role various factors play in the strengthening or weakening of the storm. In particular this study will examine correlations at peak intensity with zero time lag, and correlations of storm strength with variables frequently used in data assimilation schemes. Thus, this study ultimately looks at what factors present at the time of peak intensity contributed to the rapid intensification of Humberto.

Correlations between 850-hPa potential vorticity and surface mixing ratio and 850-hPa potential vorticity and surface temperature will be looked at first (Fig. 4.1). Correlation of 850-hPa potential vorticity with surface mixing ratio shows two broad

areas of moderate to strong correlation. One area of moderate positive correlation stretches northeast from the storm center and then turns south, forming a comma shaped pattern. Values over this broad area are greater than 0.2, with many areas reaching values of 0.35 and some areas reaching values of 0.5. In this study, correlations greater than 0.3 have 90% significance. Conversely, an area of negative correlation exists on the western side of the storm center and curls around the southeast side of the storm. South of the storm there are several areas of correlation values less than -0.35. West and northwest of the storm, however, these values increase in magnitude to greater than -0.65. As Hurricane Humberto rapidly strengthened, it benefited from the warm, moist tropical air located to its south and east. This moist air was drawn northwards and into the center of the storm, leading to the positive correlations on the eastern and southern side of the storm. Conversely, the rapid strengthening of Humberto led to increased northerly winds on the west side of the storm. These strong winds drew cooler, drier air from behind the cold front southward. Thus, high negative correlations are seen on the northwest side. In addition, lower temperatures also being drawn down the western side of Humberto.

In conjunction with those findings, the correlation structure of 850 hPa potential vorticity with surface temperature shows negative correlation areas northwest of the storm, an area of positive correlation centered directly under the storm, and a tongue of negative correlation extending from south of the storm, to just east of the storm. Clearly, the area of positive correlation directly under the storm indicates that warm air directly under the storm fuels the intensification process primarily through wind-induced surface

heat exchange (WISHE). The large area of negative correlation wrapping around the storm can be explained in much the same way as the negative correlation of mixing ratio. As the storm strengthens (potential vorticity increases), cold air from behind the cold front is pulled southward and wrapped around the storm.

Next, in order to better understand the upper level environment surrounding Humberto, the correlation of 850 hPa potential vorticity with 200 hPa zonal and meridional winds is explored (Fig. 4.2). When inspecting the correlation with 200 hPa zonal winds, an area of moderate negative correlation exists southwest of the storm and moderate positive correlation to the southeast of the storm. Together, these correlation fields indicate that a stronger storm is associated with upper level divergence south of the storm. When this is compared to a map of correlation of 850 hPa with reflectivity, a stronger storm is also correlated with higher reflectivity to the south of the storm. Thus, divergence in the upper level winds aid in the "venting" of convection, leading to a strengthened storm. The same general phenomenon can be seen in the correlation of 850 hPa potential vorticity with 200 hPa meridional winds. Although the correlation structures are not as strong as in the zonal wind case, there is still an area of divergence east and south of the storm center aiding in the ventilation of the storm at upper levels.

For the purposes of impact on data assimilation, the dynamics of Hurricane Humberto have been explored through correlation structures of variables important to assimilation schemes. It is found that the variables important to data assimilation indeed played an important role in the intensification of Hurricane Humberto. In addition, Hurricane Humberto strongly influenced the surrounding environment. Upper air data

contains significant amounts of data about storm strength and should be assimilated when possible. Venting of the storm through divergent upper level winds was associated with the development of strong convection in all quadrants of the mature Hurricane Humberto.

5. IMPACT OF STORM-CENTERED ASSIMILATION SCHEME

Significant amounts of non-Gaussianity have been found in the WRF Humberto ensemble distribution and these have been shown to potentially affect correlation structures. Because of this, it becomes even more important to find techniques that can improve the EnKF scheme. Assimilating observations in a Storm-centered (storm-centered) frame is one potential method for improvement that is independent of the issues related to non-Gaussianity. In this section, the benefits and drawbacks of a Storm-centered assimilation window will be characterized.

This study theorizes that assimilating observations in a Storm-centered frame may prove especially effective for intense mesoscale storm systems such as MCVs, intense extratropical lows, and tropical cyclones by centering the storm in the spatial assimilation window. In most assimilation schemes, including the EnKF, the assimilation domain remains fixed while the weather features and storm system move within it. When an ensemble forecast is used, however, ensemble spread of a storm center's position can be significant (up to 80 km in this study). Because of this, smaller scale features of the system can be smeared out when the ensemble average is taken and used in assimilation updates (Eq. 1.1). In the setup for the Storm-centered coordinate system, each ensemble member is shifted so that the minimum surface pressure is co-located with the ensemble mean position. Once each ensemble members' storm center has been co-located with the ensemble mean minimum pressure location, the new ensemble mean intensity is calculated, covariance matrices are constructed, and the

EnKF update is performed in accordance with equations 1.1-1.3. Position updates will still be performed in the Eulerian frame. After the ensemble mean is updated, individual ensemble members are updated using

$$\mathbf{x}_n^a - \hat{\mathbf{x}}_a = [\mathbf{I} - \beta(\hat{c}/\hat{d})\mathbf{H}](\mathbf{x}_n^f - \hat{\mathbf{x}}^f) \qquad (5.1)$$

and then translated back to their original position on the Eulerian grid using the same linear transformation as in the first step. For this study, β is set to one. In equation 5.1, the term on the left hand side, $\mathbf{x}_n^a - \hat{\mathbf{x}}_a$, is the analysis perturbation, or the difference between the analysis of ensemble member n and the ensemble mean, **I** is the identity matrix, $\hat{c}$ is the forecast covariance of the state variable with observed variable, $\hat{d}$ is the sum of the forecast covariance of the observation variable and the observation error (variance), and **H** is a transformation matrix that maps model space to observation space. Finally, $(\mathbf{x}_n^f - \hat{\mathbf{x}}^f)$ is the forecast perturbation of ensemble member n. Thus, in a simple example, the analysis perturbation is the forecast perturbation reduced by one minus the Kalman gain matrix. Once this calculation has been performed, the updated ensemble members can be used as the new background state and the process repeated with a new set of observations.

Results from the two methods will be compared, looking at the amount of error reduction from each method. Assimilation of an intensity observation only is a straightforward experiment from which comparisons can be made. Ensemble member eleven is sufficiently accurate in describing the true evolution of Humberto to be chosen as the "truth" member. The pressure field for this truth member is shown in Figure 5.1. From the truth member, a simulated pressure observation is taken at the gridpoint with

minimum pressure and random Gaussian noise with $\sigma^2 = 1$ millibar is added. The observation is then treated as if it is co-located with the ensemble mean and assimilated in both the Eulerian and Storm-centered schemes.

Comparing the initial Storm-centered and Eulerian ensemble means as shown in Figure 5.2, the most notable difference is the smaller, more compact, storm center noted in the Storm-centered frame. The Storm-centered agrees more closely with the truth and also agrees well with the idea that the Eulerian mean tends to smear mesoscale features of the mesoscale system. Using the EnKF scheme outlined in the introduction of this manuscript, the posterior pressure distributions shown in Figure 5.3 are obtained. These estimates appear similar, with the Storm-centered update being slightly more symmetrical and the Eulerian update being slightly weaker on the west side. In contrast, the Storm-centered and Eulerian update have different patterns in the increment. Unlike the Eulerian increment, which has a dipole pattern, the Storm-centered update is extremely symmetrical around the storm center. This indicates that the Storm-centered update focuses on an area concentrated around the storm center while the Eulerian update is affected by the different locations of the storm center.

To fully understand the quality of the update in each frame, the performance of the Storm-centered and Eulerian update after transforming the ensemble members from the Storm-centered update back to the Eulerian frame is examined. While updating intensity in the storm-centered frame, position is updated with a position observation taken from the truth member in the Eulerian frame. Thus, the ensemble mean from the intensity update in the storm-centered frame is then moved to the location indicated by

the position update. The post-transformation Storm-centered update ensemble mean remains fairly symmetrical, but is stronger than the original ensemble mean. Unlike the Eulerian update, the transformed Storm-centered update shows no erosion on the western side of the storm and presents a very symmetrical ensemble mean.

To quantitatively assess the quality of the two methods of assimilating pressure observations, a new measure of error is introduced. This measure can be written as

$$E_N = \frac{|E_{prior}| - |E_{post}|}{|\overline{E_{prior}}|} \qquad (5.2)$$

where E_N is the normalized error, E_{prior} is the error in the prior estimate, E_{post} is the error in the posterior estimate, and $\overline{E_{prior}}$ is the error in the prior estimate averaged over the entire domain. Error is measured by comparing the ensemble mean to the chosen truth member. E_N is calculated at each gridpoint. By using this measure of error instead of simply measuring the error as the magnitude of the perturbation from truth (in mb), errors of different magnitudes can be compared on equal footing. Thus, a value of zero by this measure indicates no change in the error from the prior to posterior estimate, large positive values indicate large decreases in magnitude of the error, and large negative values indicate large increases in the magnitude of the error.

Looking at this measure of performance (Fig 5.4), and the associated analyses it can be seen that both the Storm-centered and Eulerian methods have similar error reduction properties, but different analyses. Firstly, the Eulerian update produces significant improvement in a narrow band on the northeast side of the storm (the area between the ensemble mean and truth) while there is a larger area of negative E_N to the

southwest. The storm-centered scheme shows a similar pattern, but magnitudes, both positive and negative, that are slightly lower.

Overall, then, both methods reduce error in the analysis. The main difference comes in the structure of the analyses. In the Eulerian scheme analysis, the pressure field looks distorted, with higher pressures on the southwest side of the storm and an elongation of the storm from northwest to southeast. In contrast, the storm-centered update looks very symmetrical and agrees more closely with the truth member of the ensemble (Fig. 5.1). While the Eulerian method provides the highest magnitude normalized error reduction in limited areas, it does not provide as accurate of a picture of the structure of the storm. Each type of assimilation method produces more accurate initial conditions for a model; however, the Storm-centered technique presents a new and exciting alternative for an EnKF scheme that does not distort the shape and orientation of the storm.

6. SUMMARY AND CONCLUSIONS

A study of current data assimilation methods, their strengths, and their weaknesses has been outlined. The main focus has been a characterization of ensemble Gaussianity for a L95 Lorenz model and a rapidly intensifying tropical cyclone and the potential impact non-Gaussianity may have on the quality of the EnKF. In addition, an alternative, Storm-centered, or storm-centered, method of performing the EnKF update has been explored. Finally, correlation structures have been used to study the underlying dynamics leading to the rapid intensification of Hurricane Humberto. It has been shown that non-Gaussianity is present in the ensemble distribution and that the variables important in the EnKF scheme were important in the rapid intensification of Humberto.

While ensemble non-Gaussianity presents a unique statistical problem to be resolved in the theoretical world of atmospheric science research, it also presents problems for operational forecasting. Investigating the size of an ensemble necessary to retain Gaussianity may be an important question to find an answer to, but a more pressing question, in this day of limited computational resources, might be how to address the non-Gaussianity present in current ensembles. Performance of the EnKF depends on solving this non-Gaussianity problem since the EnKF is a linear theory.

In this manuscript, the initial questions surrounding these problems are formulated, and initial findings and steps to the solutions are presented. Non-Gaussianity has been firmly established in the ensemble distribution for the WRF forecast of Hurricane Humberto. To demonstrate the effect that non-Gaussianity can

have on correlation structures data transformations were performed on randomly generated Gaussian distributions. Correlations between these transformed data sets were shown to change, sometimes significantly, in a number of cases. Thus, it can be inferred that in at least a number of cases, the non-Gaussianity of the ensemble distribution affects the correlation structures used in the EnKF update.

To address the problems presented by ensemble non-Gaussianity, a storm-centered, Storm-centered assimilation method has been tested. This potential framework for performing data assimilation presents a method to potentially counteract the implications of the non-Gaussianity. By making the mesoscale system (Hurricane Humberto in this case) the center of the assimilation window, ensemble averages tend to be smoothed less. Because of this more accurate depiction of the mesoscale system, correlation structures near the storm will be more accurately calculated. In turn, more accurate correlation structures will lead to a more accurate EnKF update. However, this method presents a new set of problems. Although error is more greatly reduced in the storm-centered update, the transformation back to the Eulerian frame introduces new error, somewhat reducing it's effectiveness. In the end, this method should be tested in a broader array of environments and circumstance to determine if the promise shown in this study is real.

Finally, some of the dynamics underlying the intensification of Hurricane Humberto were explored. In particular, the role in intensification of variables important to data assimilation was explored. One of the more interesting findings was the large amount of information transferred from upper levels (200-hPa) to lower levels (850-hPa)

as seen in the correlation of upper level zonal and meridional winds to 850-hPa potential vorticity. This finding would urge the data assimilation community to not discount assimilation of unconventional observations (upper level winds), as they may prove vital to assimilation schemes.

In the end, a view of ensemble characteristics, data assimilation methods, and the relation between dynamics and data assimilation has been presented from a new perspective. However, the conclusions and observations drawn from this data are limited by the fact that they come from a single case. For example, do other model ensembles show the same non-Gaussianity as the WRF model did in this particular case? Does the Storm-centered method of data assimilation produce similar results for other storms or non-tropical systems? Few papers have answered these questions. They are questions that must be answered to continue the march towards better forecasting. It will not only further improve our understanding of numerical weather prediction and tropical cyclones, but it will save money and lives.

REFERENCES

Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev*., **129**, 2884-2903.

Blake, E. S., 2007: Tropical cyclone report Hurricane Humberto. [http://www.nhc.noaa.gov/2007atlan.shtml]

Elsberry, R.L., Lambert, T.D.B. & Boothe, M.A., 2007: Accuracy of Atlantic and eastern North Pacific tropical cyclone intensity forecast guidance. *Wea. Forecasting*, **22**, 747-762.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**, 10143-10162.

Franklin, J. L., 2005: 2004 National Hurricane Center verification report. *Preprints*, 57th Interdepartmental Hurricane Conference, Miami, FL. [Updates are available on the official National Hurricane Center website: National Hurricane Center (NHC), www.nhc.noaa.gov/verification]

Grell, G.A. and D. Devenyi, 2002: A generalized approach to parameterizing convection combining ensemble and data assimilation techniques. *Geophys. Res. Let*., **29**, 10.1029/2002GL015311.

Hawbliztel, D., F. Zhang, Z. Meng and C. A. Davis, 2007: Probabilistic evaluation of the dynamics and predictability of mesoscale convective vortex event of 10-13 June 2003. *Mon. Wea. Rev*., **135**, 1544-1563.

Hong, S-.Y. , J. Dudhia, and S.-H. Chen, 2004: A revised approach to ice-microphysical processes for the bulk parameterization of cloud and precipitation. *Mon. Wea. Rev*., **132**, 103-120.

Houtekamer, P.L., Mitchell, H.L., Pellerin, G., Buehner, M., Charron, M., Spacek, L., Hansen, B., 2005: Atmospheric data assimilation with an ensemble kalman filter: Results with real observations. *Mon. Wea. Rev*., **133**, 604-620.

Houtekamer, P. L. and Mitchell, H.L., 2001: A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev*., **129**, 123-137.

Houtekamer, P.L. and Mitchell, H., 1998: Data assimilation using an ensemble kalman filter technique. *Mon. Wea. Rev*., **126**, 796-811.

Jazwinski, A.H., 1970: Stochastic processes and filtering theory. New York: Academic Press.

Kalman, R. E., 1960: A new approach to linear filtering and prediction problems. *J. Basic Eng.*, **82**D, 35-45.

Lawson, W., Hansen, J., 2005: Alignment error models and ensemble-based data assimilation. *Mon. Wea. Rev.*, **133**, 1687-1709.

Meng, Z, and F. Zhang, 2007: Test of an ensemble-Kalman filter for mesoscale and regional-scale data assimilation. Part II: Imperfect-model experiments. *Mon. Wea. Rev*., **135**, 1403-1423.

Meng, Z, and F. Zhang, 2008a: Test of an ensemble-Kalman filter for mesoscale and regional-scale data assimilation. Part III: Comparison with 3Dvar in a real-data case study. *Mon. Wea. Rev*., 136, 522-540.

Meng, Z, and F. Zhang, 2008b: Test of an ensemble-Kalman filter for mesoscale and regional-scale data assimilation. Part IV: Performance over a warm-season month of June 2003. *Mon. Wea. Rev*., in press.

Noh, Y., W.-G. Cheon, and S.-Y. Hong, 2003: Improvement of the K-profile model for the planetary boundary layer based on large eddy simulation data. *Bound.-Layer Meteor*., **107**, 401-427.

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, W. Wang, and J. G. Powers, 2005: A description of the advanced research WRF version 2. *NCAR Technical Note*, NCAR/TN-468+STR.

Snyder, C., and F. Zhang, 2003: Tests of an ensemble Kalman filter for convective-scale data assimilation. *Mon. Wea. Rev*., *131*, 1663-1677.

Zhang, F., 2005: Dynamics and structure of mesoscale error covariance of a winter cyclone estimated through short-range ensemble forecasts. *Mon. Wea. Rev*., **133**, 2876-2893.

Zhang, F., Z. Meng and A. Aksoy, 2006: Test of an ensemble-Kalman filter for mesoscale and regional-scale data assimilation. Part I: Perfect-model experiments. *Mon. Wea. Rev*., **134**, 722-736.

Zhang, F., Weng, Y., Sippel, J., Meng, Z., Bishop, C., 2009: Cloud-resolving hurricane initialization and prediction through assimilation of doppler radar observations with an ensemble kalman filter: Humberto (2007). *Mon. Wea. Rev.* (in review).

APPENDIX A

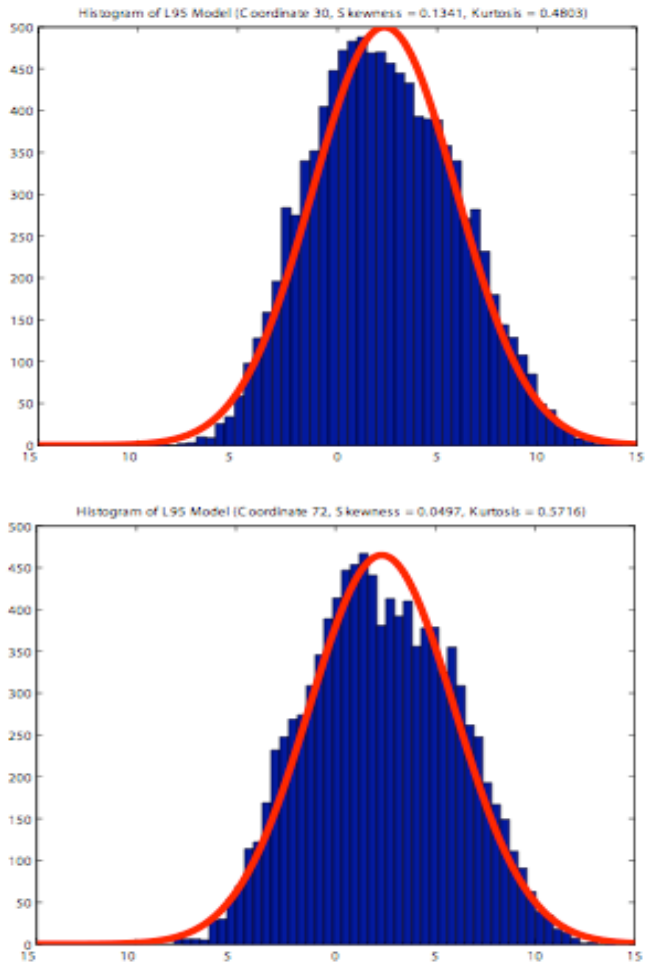

Figure 2.1: Sample ensemble distributions of L95 Lorenz model. Red line indicates Gaussian overlay.

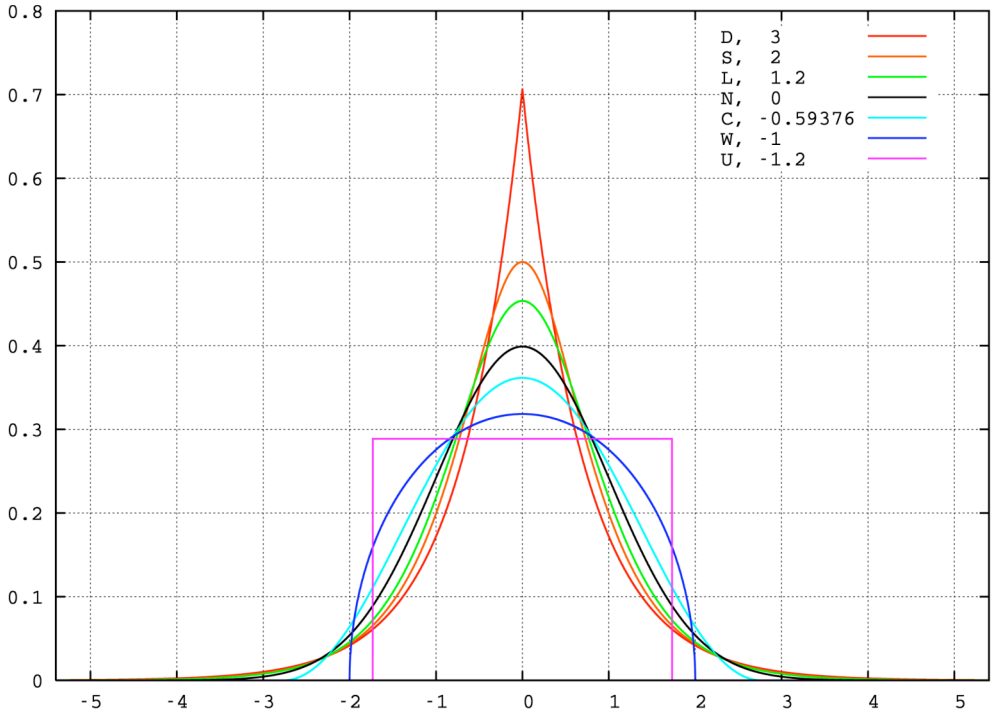

Figure 2.2: Sample distributions of unimodal distributions. Kurtosis values range from -1.2 to 3.0.

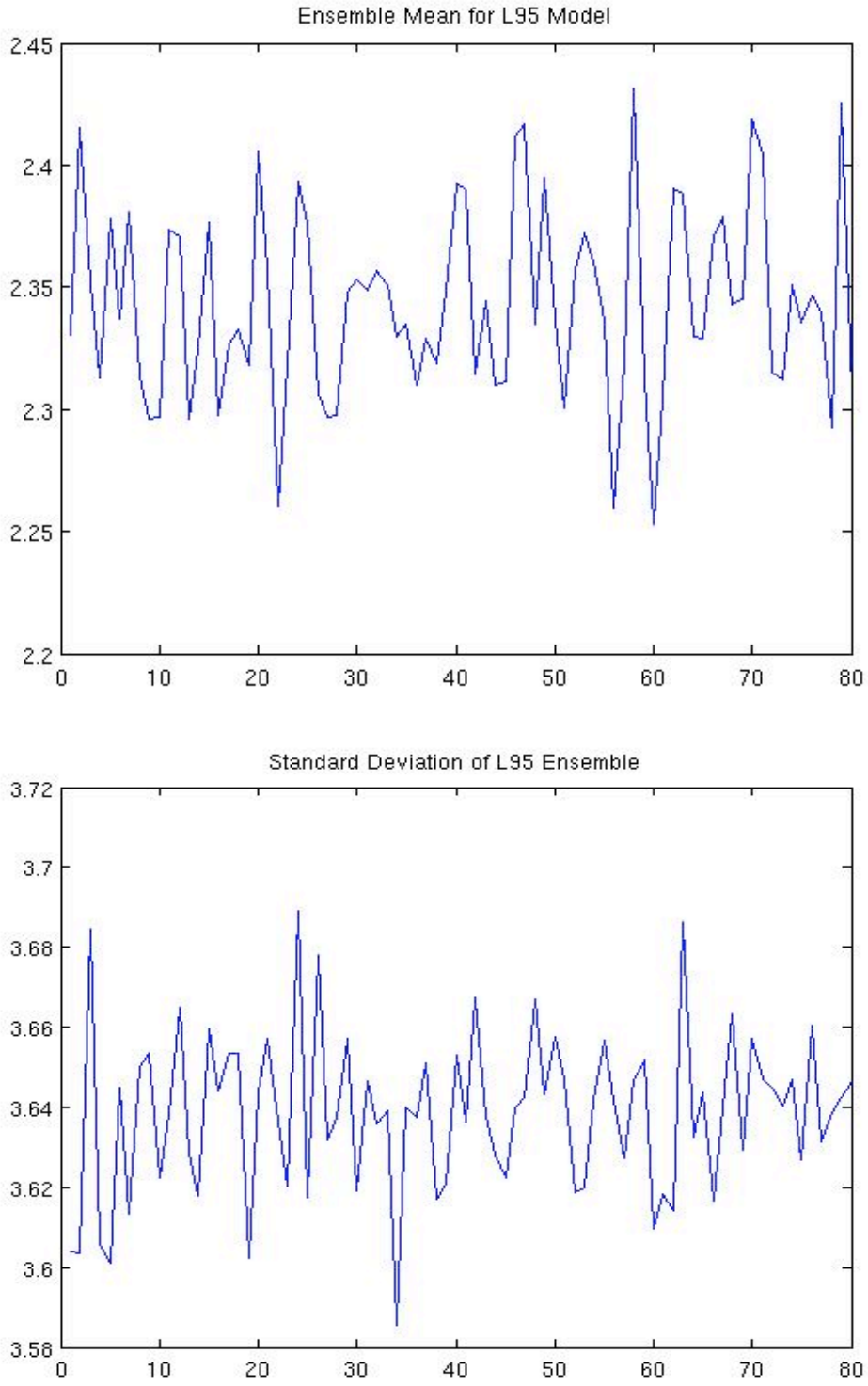

Figure 2.3: Ensemble mean (top) and standard deviation (bottom) of L95 Lorenz model.

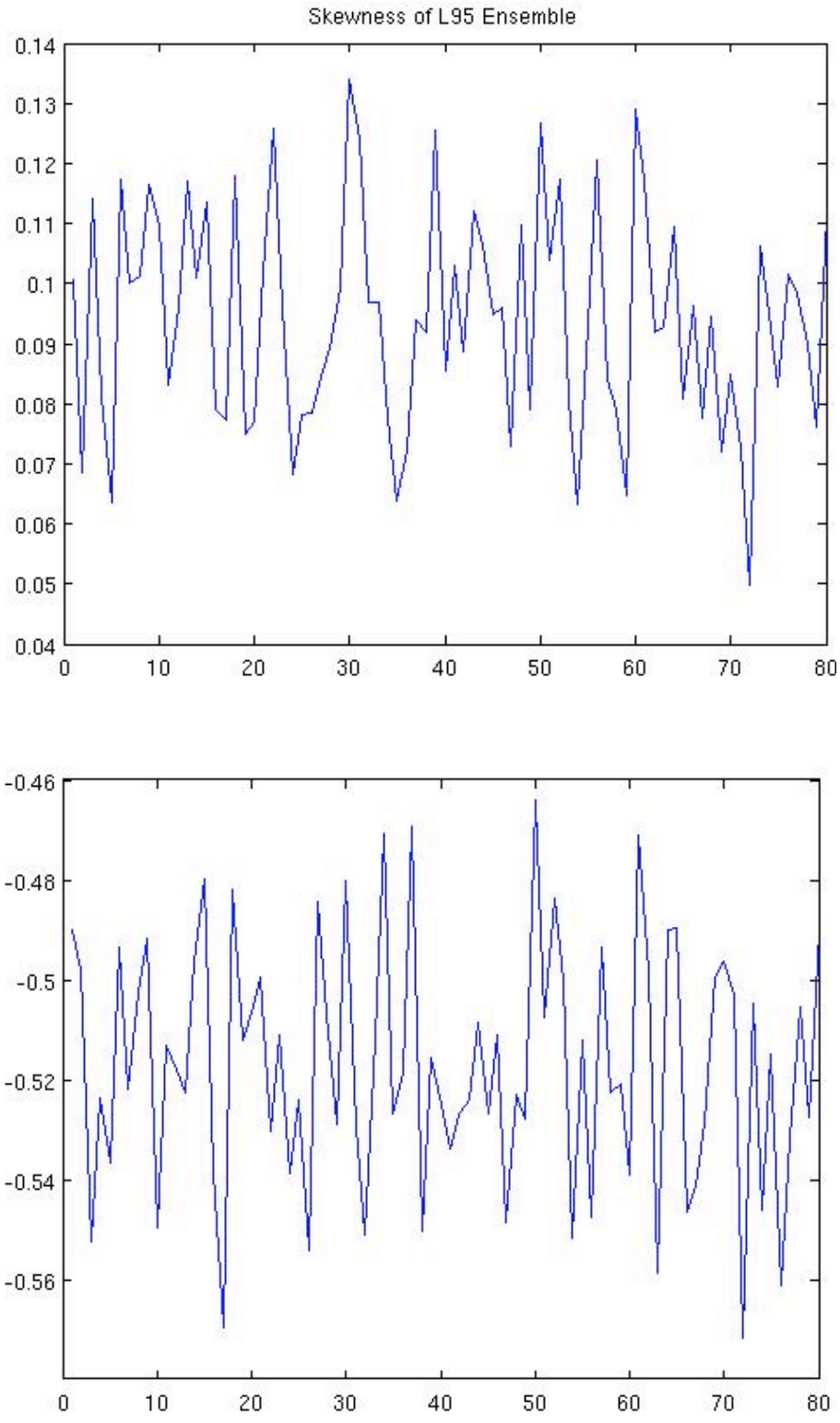

Figure 2.4: Ensemble skewness (top) and kurtosis (bottom) of L95 Lorenz model.

Ensemble Size	N = 10,000	N = 30
Mean STD	3.6386	3.5943
Max STD	3.6892	4.5561
Min STD	3.6386	2.7174
STD(STD)	0.0203	0.3905
Mean Skew	0.0940	0.1357
Max Skew	0.1314	1.4416
Min Skew	0.0497	-1.1651
STD(Skew)	0.0184	0.4961
Mean Kurtosis	-0.5172	3.4395
Max Kurtosis	-0.4640	7.2326
Min Kurtosis	-0.5716	1.7024
STD(Kurtosis)	0.0189	1.1686
Mean(Ensemble Mean)	2.3433	2.3849
Max Ensemble Mean	2.4312	4.0473
Min Ensemble Mean	2.2528	0.8714
STD Ensemble Mean	0.0397	0.6982

Table 2.1: Effect of ensemble size on Gaussianity. Left hand column shows descriptive statistics for 10,000 member ensemble. Right hand column shows descriptive statistics for 30 member ensemble.

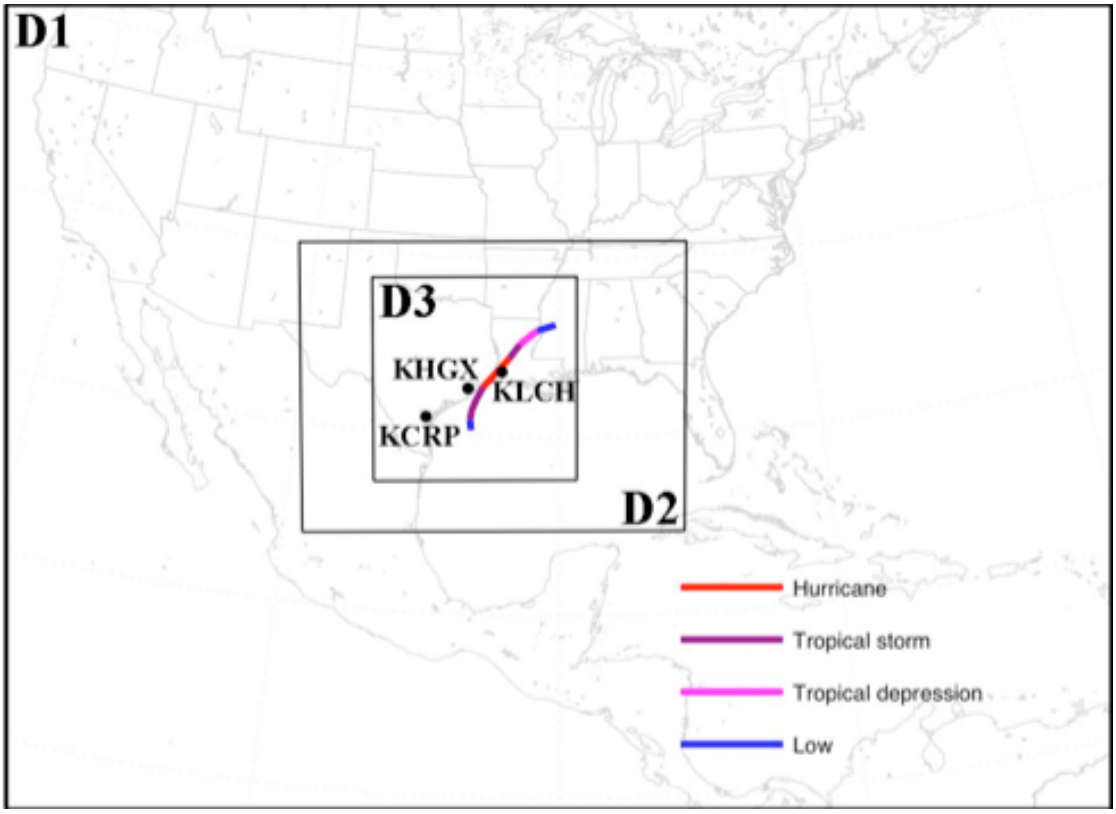

Figure 3.1: Configuration of the WRF model domains 1, 2, and 3. Horizontal grid spacings of 40.5, 13.5, and 4.5 km, respectively. Also depicted are NHC best-track estimate of Humberto with intensity color-coded and three WSR-88D radar locations.

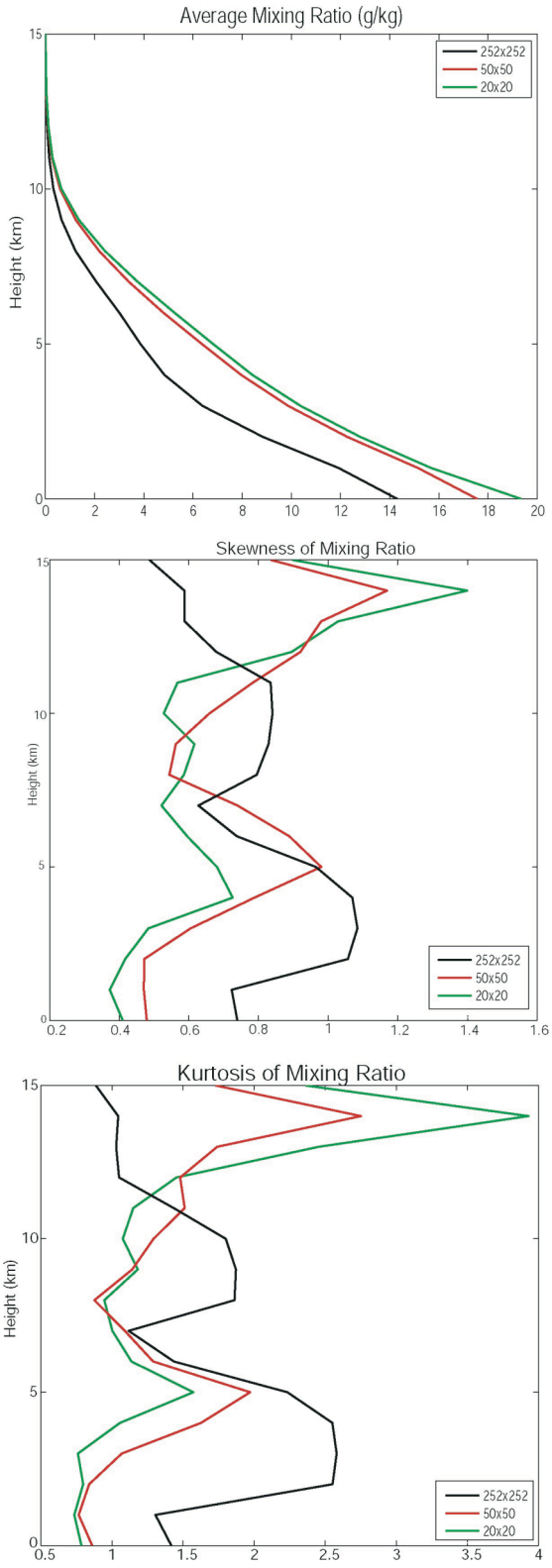

Figure 3.2: Ensemble mean mixing ratio (top) (in g/kg), skewness (middle), and kurtosis (bottom). Black line indicates area average over 252x252 gridpoints, red line indicates average over 50x50 gridpoints, and green line indicates average over 20x20 gridpoints.

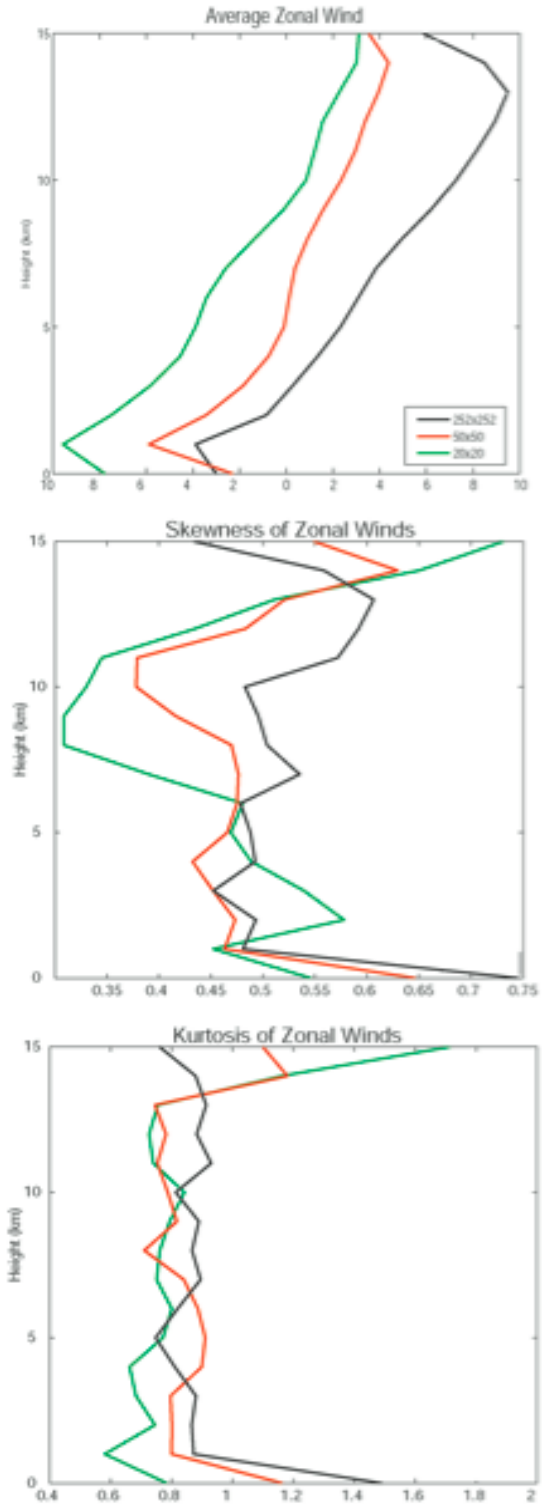

Figure 3.3: Ensemble mean zonal winds (top) (in m/s), skewness (middle), and kurtosis (bottom). Black line indicates area average over 252x252 gridpoints, red line indicates average over 50x50 gridpoints, and green line indicates average over 20x20 gridpoints.

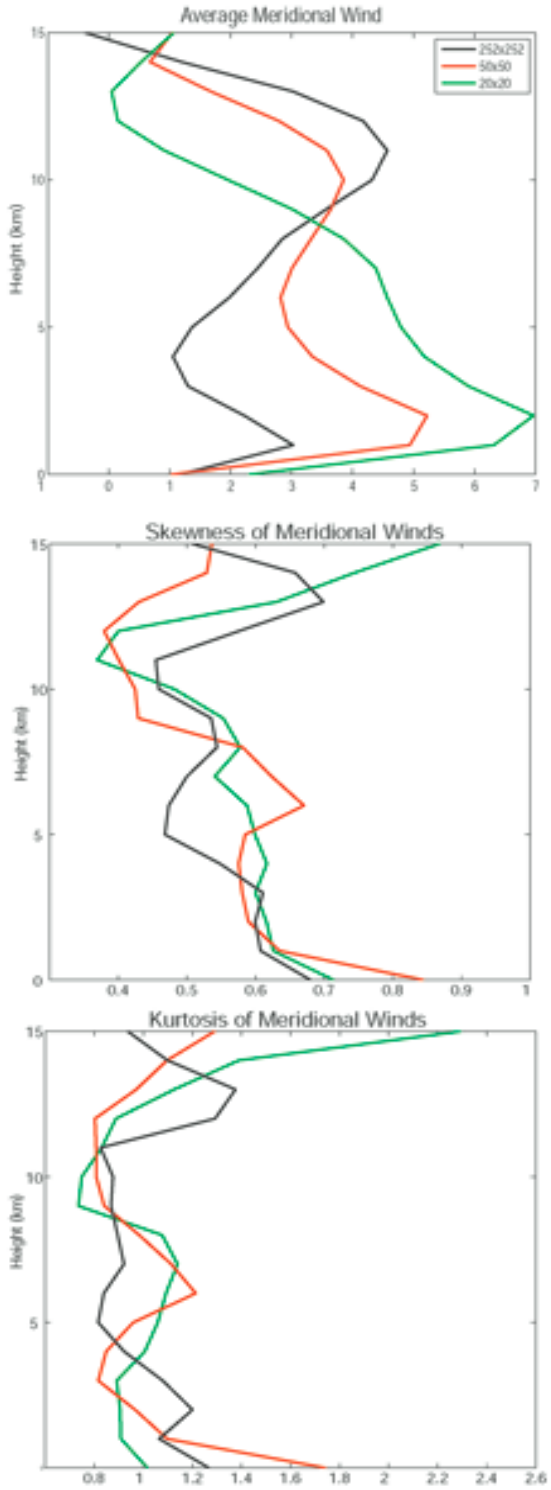

Figure 3.4: Ensemble mean meridional winds (top) (in m/s), skewness (middle), and kurtosis (bottom). Black line indicates area average over 252x252 gridpoints, red line indicates average over 50x50 gridpoints, and green line indicates average over 20x20 gridpoints.

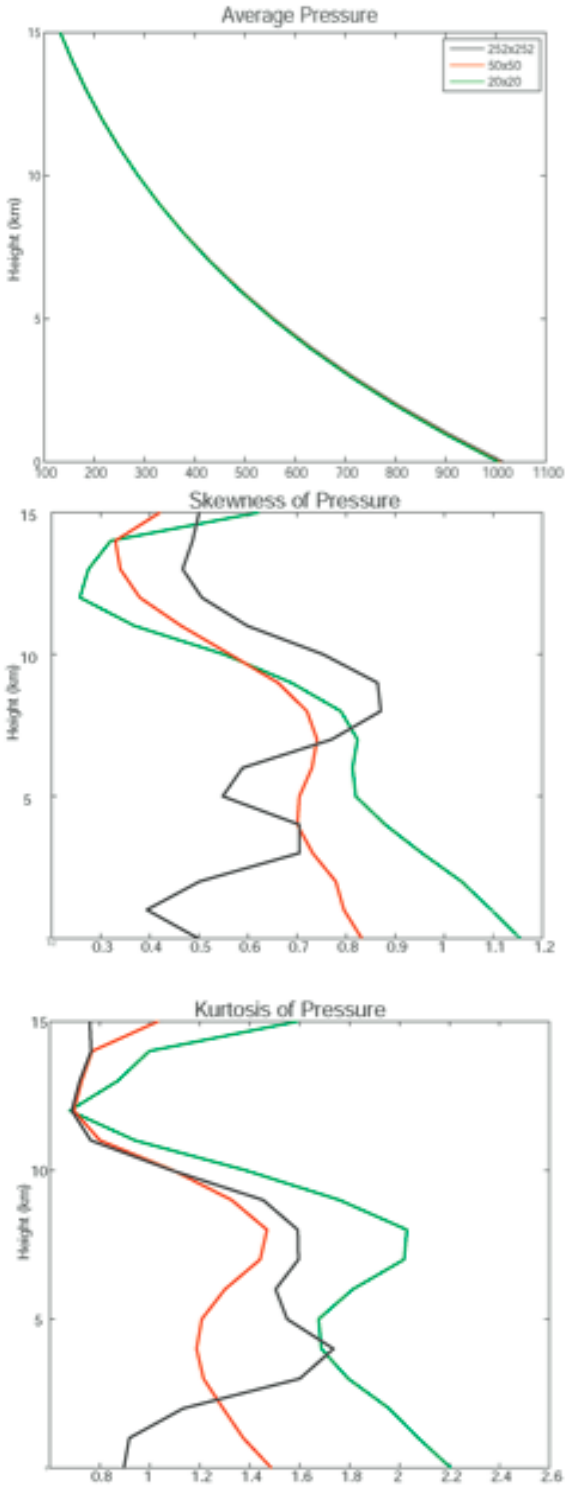

Figure 3.5: Ensemble mean pressure (top) (in mb), skewness (middle), and kurtosis (bottom). Black line indicates area average over 252x252 gridpoints, red line indicates average over 50x50 gridpoints, and green line indicates average over 20x20 gridpoints.

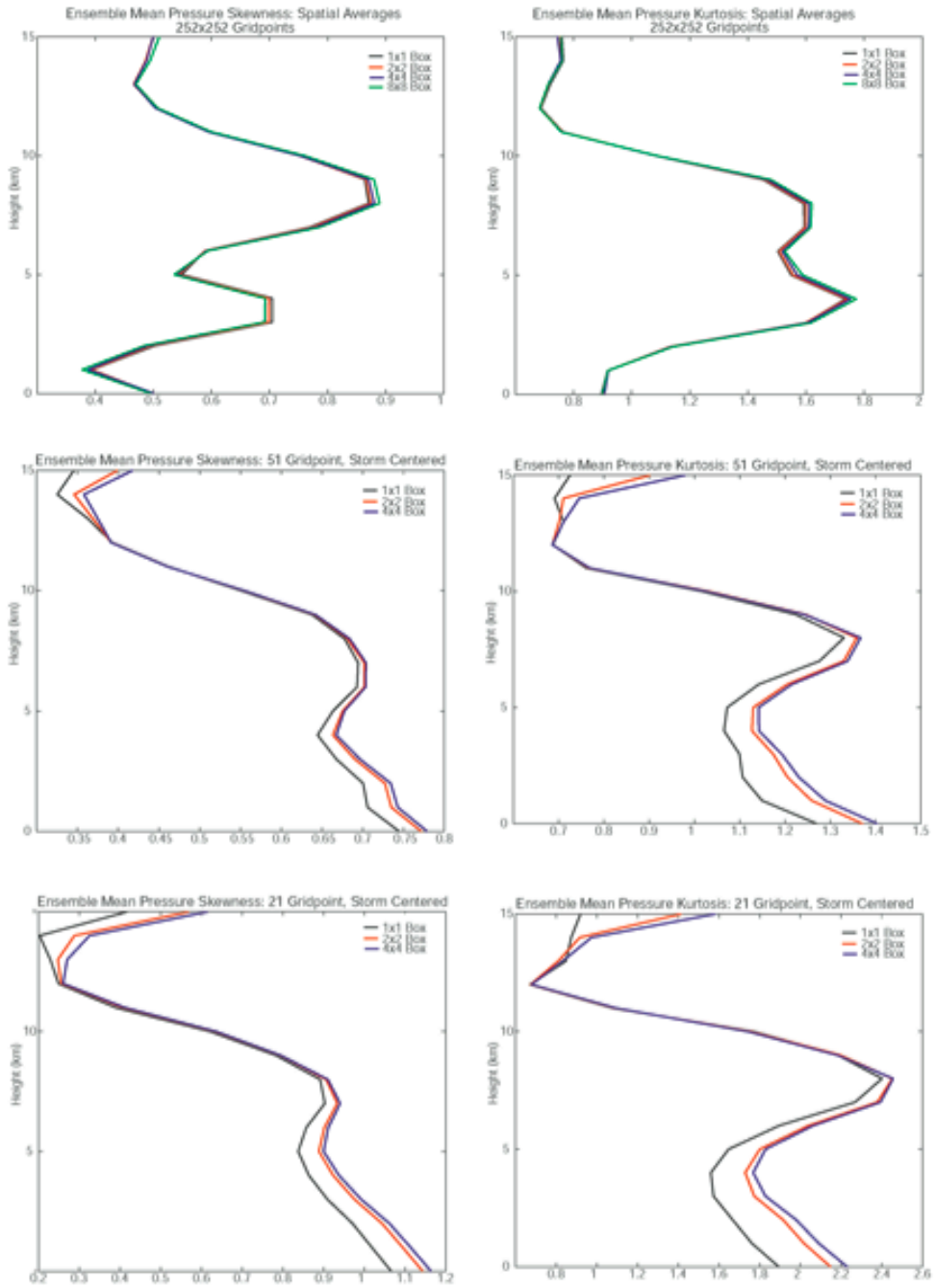

Figure 3.6: Effect of area averaging on Gaussianity of pressure. Left hand column skewness of pressure and right hand column shows kurtosis. Top row averages over entire 252x252 gridpoint domain, middle column averages over 50x50 gridpoint domain, and bottom row averages over 20x20 domain. In top row, black line indicates no averaging, red line averaging over 2x2 box, blue a 4x4 box, and green an 8x8 box. In the bottom two rows, blue indicates no averaging, red indicates averaging over a 2x2 box, and black indicates averaging over a 4x4 box.

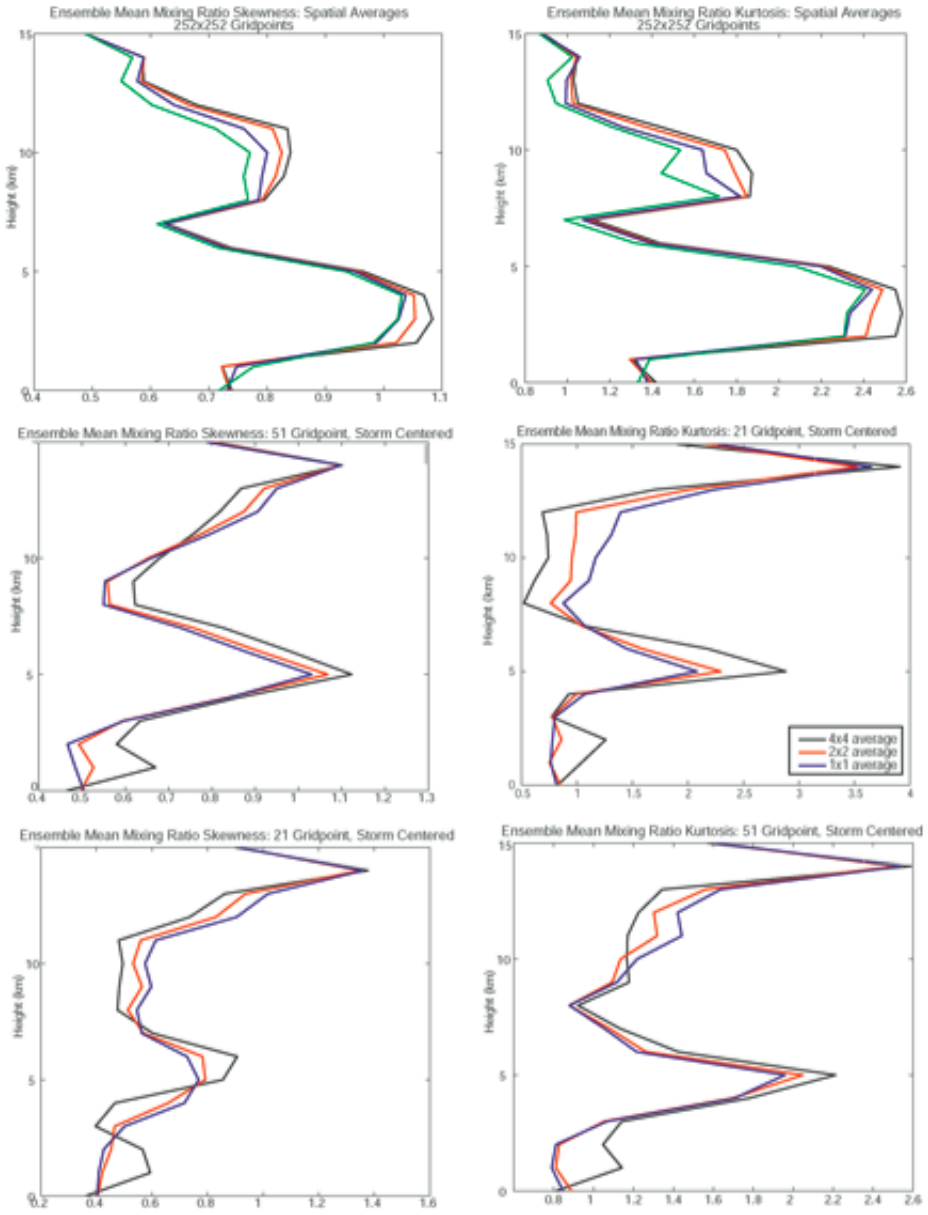

Figure 3.7: Effect of area averaging on Gaussianity of mixing ratio. Left hand column skewness of mixing ratio and right hand column shows kurtosis. Top row averages over entire 252x252 gridpoint domain, middle column averages over 50x50 gridpoint domain, and bottom row averages over 20x20 domain. In top row, black line indicates no averaging, red line averaging over 2x2 box, blue a 4x4 box, and green an 8x8 box. In the bottom two rows, blue indicates no averaging, red indicates averaging over a 2x2 box, and black indicates averaging over a 4x4 box.

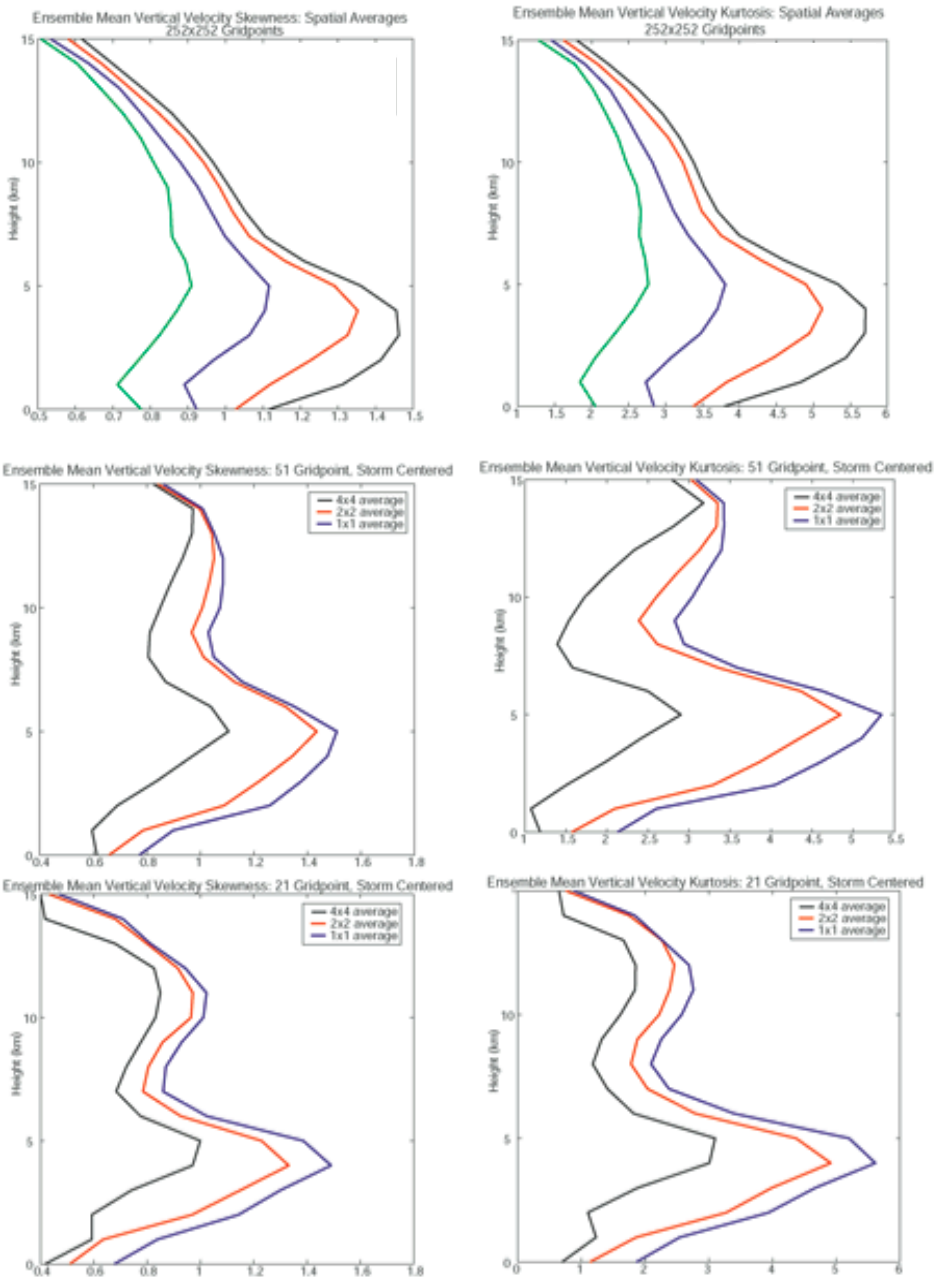

Figure 3.8: Effect of area averaging on Gaussianity of vertical velocity. Left hand column shows skewness and right hand column shows kurtosis. Top row averages over entire 252x252 gridpoint domain, middle column averages over 50x50 gridpoint domain, and bottom row averages over 20x20 domain. In top row, black line indicates no averaging, red line averaging over 2x2 box, blue a 4x4 box, and green an 8x8 box. In the bottom two rows, blue indicates no averaging, red indicates averaging over a 2x2 box, and black indicates averaging over a 4x4 box.

	Gaussian Distribution	x^2 Transformation	x^3 Transformation	1/x Transformation
Mean = 0, Std. Dev. = 1				
Correlation 1	0.0076	0.0588	0.7491	0.0393
Correlation 2	N/A	0.9932	0.9849	0.0096
Skewness A	0.0492	2.7271	0.1316	21.5788
Skewness B	0.0856	3.1520	1.6469	-23.5874
Kurtosis A	0.0450	9.8026	25.3083	622.1828
Kurtosis B	0.0798	14.7053	42.2425	670.1089
Mean = 270, Std. Dev. = 1				
Correlation 1	.9984	0.9984	0.9983	-0.9985
Correlation 2	N/A	0.9984	0.9984	0.9984
Skewness A	-0.0953	-0.0842	-0.0731	0.1175
Skewness B	0.0020	0.0140	0.0140	0.0221
Kurtosis A	-0.0072	-0.0093	-0.0110	-0.0017
Kurtosis B	0.0696	0.0684	0.0677	0.0735
Mean = -15, Std. Dev. = 1				
Correlation 1	0.9965	-0.9950	0.9901	-0.9913
Correlation 2	N/A	0.9965	0.9962	0.9958
Skewness A	-0.0074	0.2559	-0.5242	-0.4624
Skewness B	-0.0347	0.2671	-0.4967	-0.4684
Kurtosis A	0.4096	0.7296	1.3749	0.5173
Kurtosis B	0.3367	0.3915	0.6687	0.9984

Table 3.1: Effect of non-Gaussian transformations on correlation structures. Correlation one corresponds to the correlation of one transformed distribution with one Gaussian distribution. Correlation 2 indicates correlation of two transformed distributions (except in case of "**Gaussian Distribution**" in which "Correlation 1" is the correlation of two Gaussian distributions).

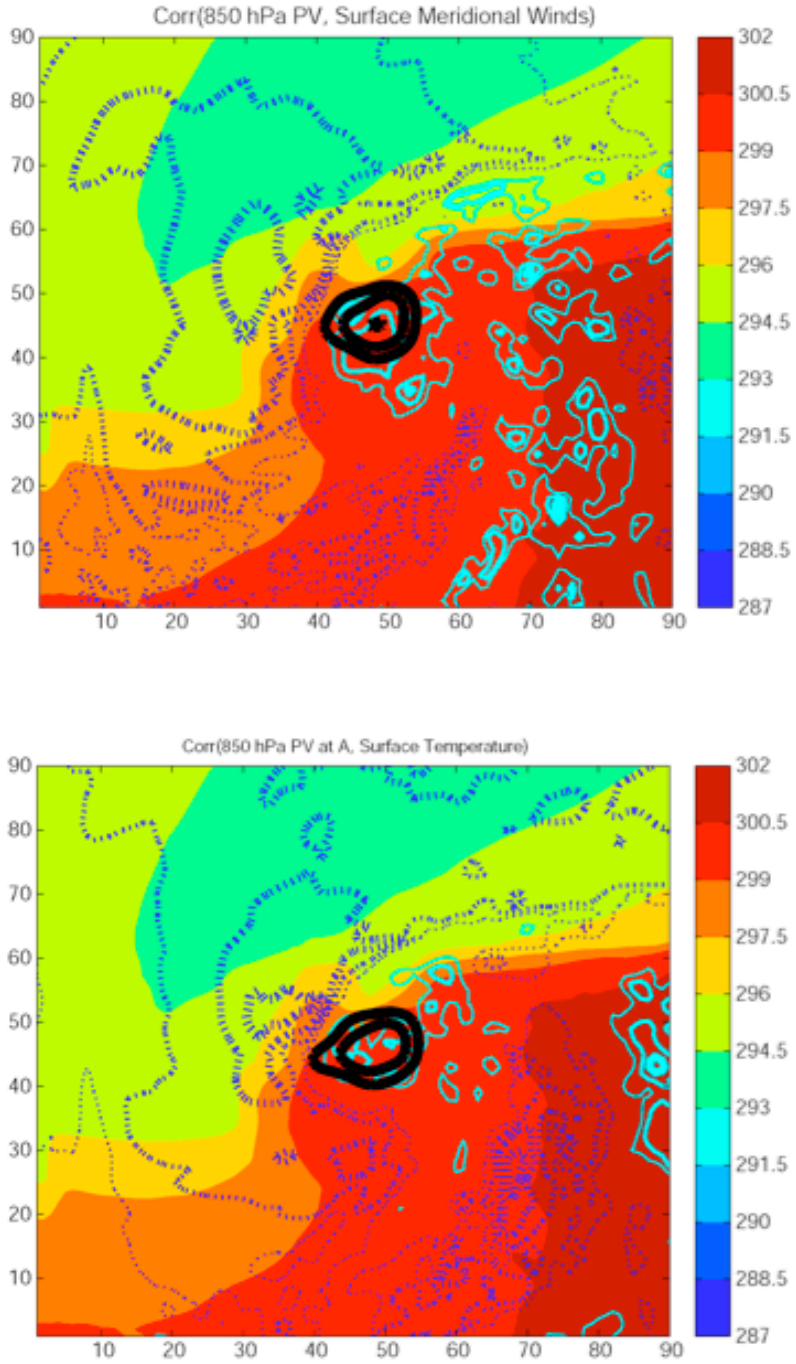

Figure 4.1: Correlation of 850 hPa potential vorticity with surface mixing ratio (top) and surface temperature (bottom). Dashed contours indicate negative correlation, solid contours indicate positive correlation. Lowest magnitude correlation value represented by thinnest line at a magnitude of 0.2. Filled contours show surface temperature in Kelvin.

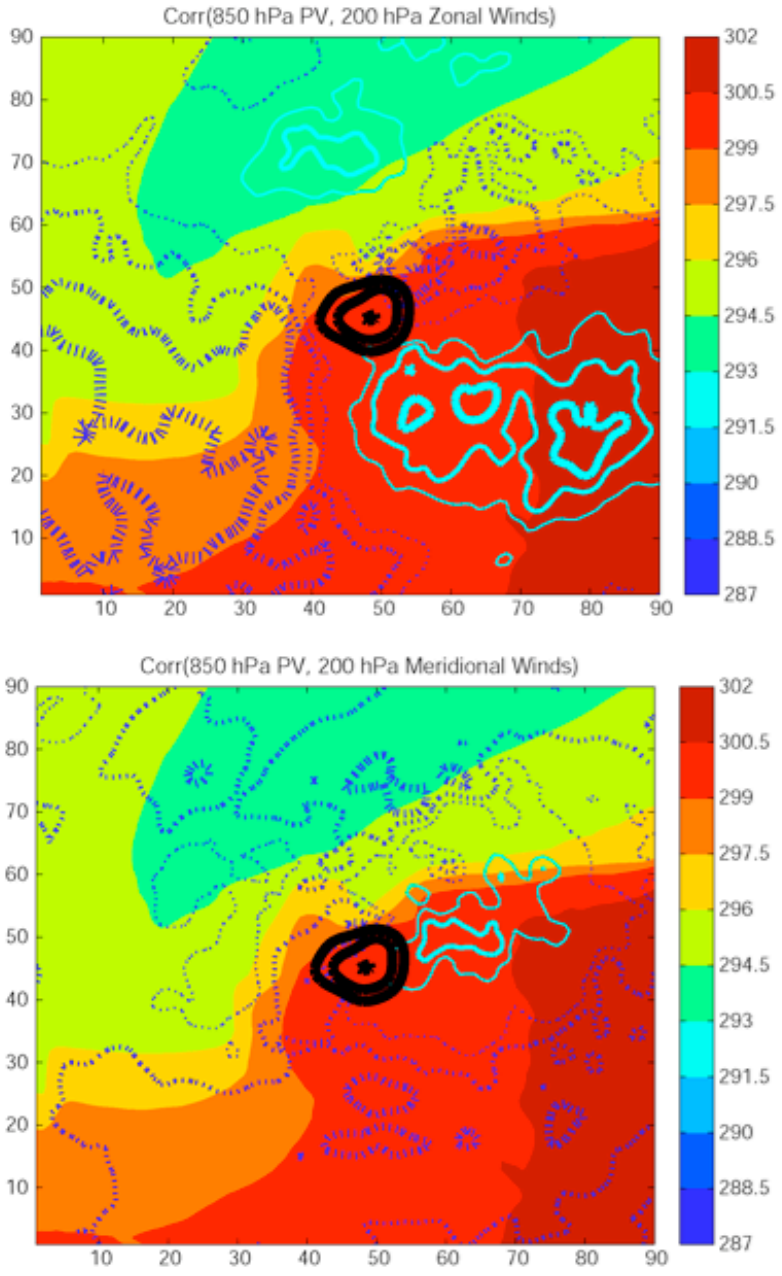

Figure 4.2: Correlation of 850 hPa potential vorticity with 200 hPa zonal wind (top) and meridional wind (bottom). Dashed contours indicate negative correlation, solid contours indicate positive correlation. Lowest magnitude correlation value represented by thinnest line at a magnitude of 0.2. Filled contours show surface temperature in Kelvin.

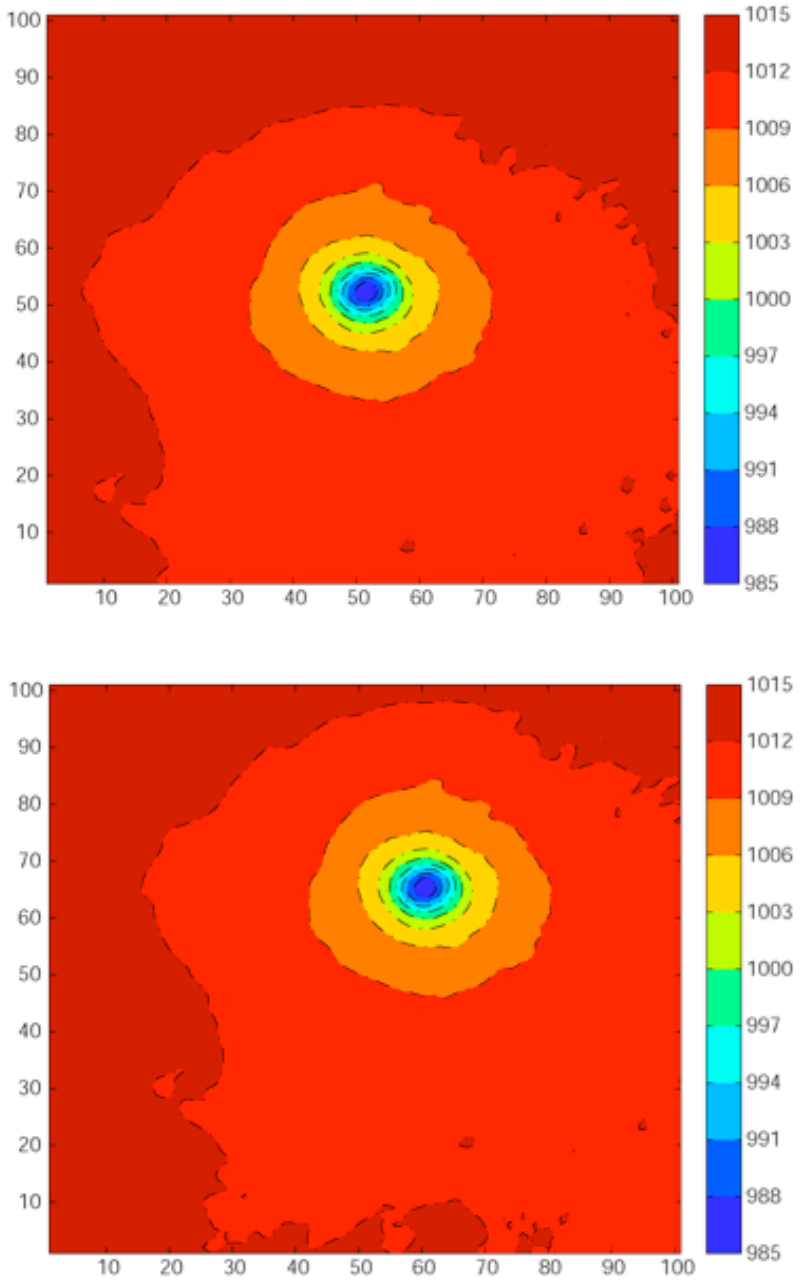

Figure 5.1: Pressure field of ensemble member 11 ("truth" member). Storm is shown in Storm-centered coordinates (top) and Eulerian coordinates (bottom). Pressure scale is in millibars.

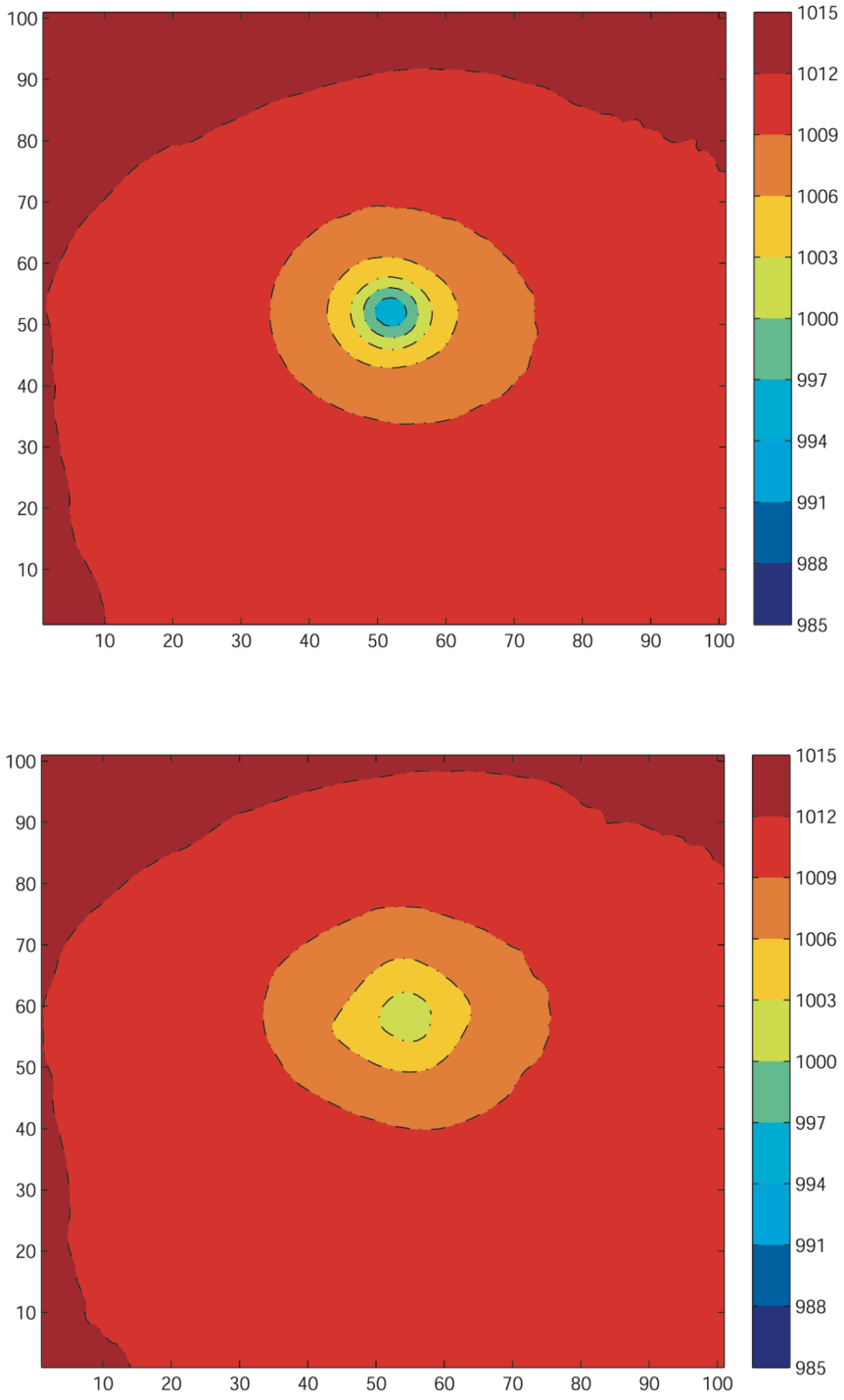

Figure 5.2: Ensemble mean pressure of storm-centered (top) and Eulerian frames (bottom). Pressure scale is in millibars.

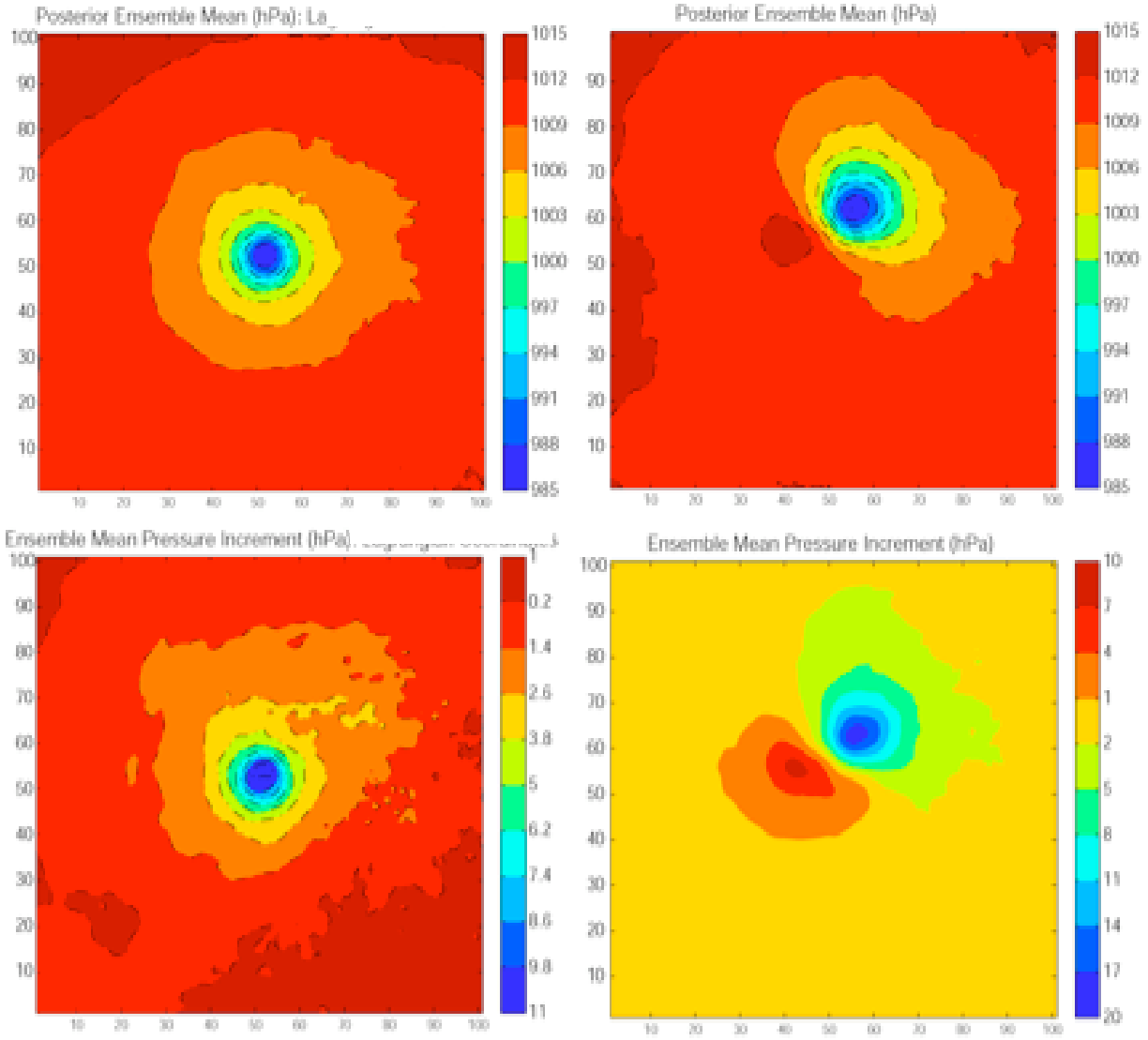

Figure 5.3: Storm-centered coordinate system (top left) and Eulerian coordinate system (top right) analyses. Analysis increment shows the difference between the posterior analysis and the prior estimate for the Storm-centered frame (bottom left) and Eulerian frame (bottom right). Pressure scale is in millibars.

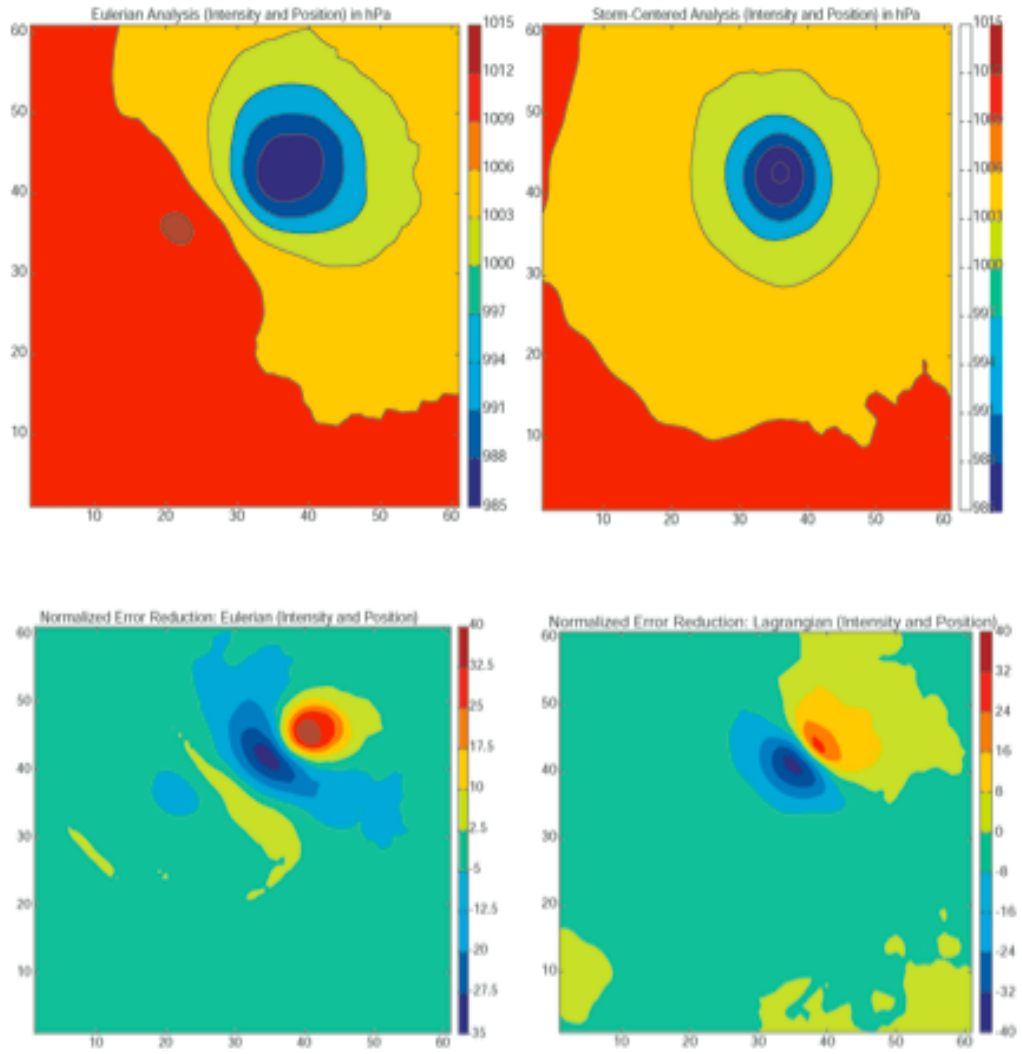

Figure 5.4: Normalized measure of error reduction. EnKF analysis in the Eulerian frame (top left), Storm-centered frame (top right). Normalized error reduction of Eulerian update (bottom left) and storm-centered frame (bottom right).

VITA

Name: Matthew Charles Rigney

Address: Department of Atmospheric Sciences
Texas A&M University
3150 TAMU
College Station, TX 77843-3150

Email Address: matt.rigney@gmail.com

Education: B.S., Physics, Rice University, 2006
M.S., Atmospheric Science, Texas A&M University, 2009